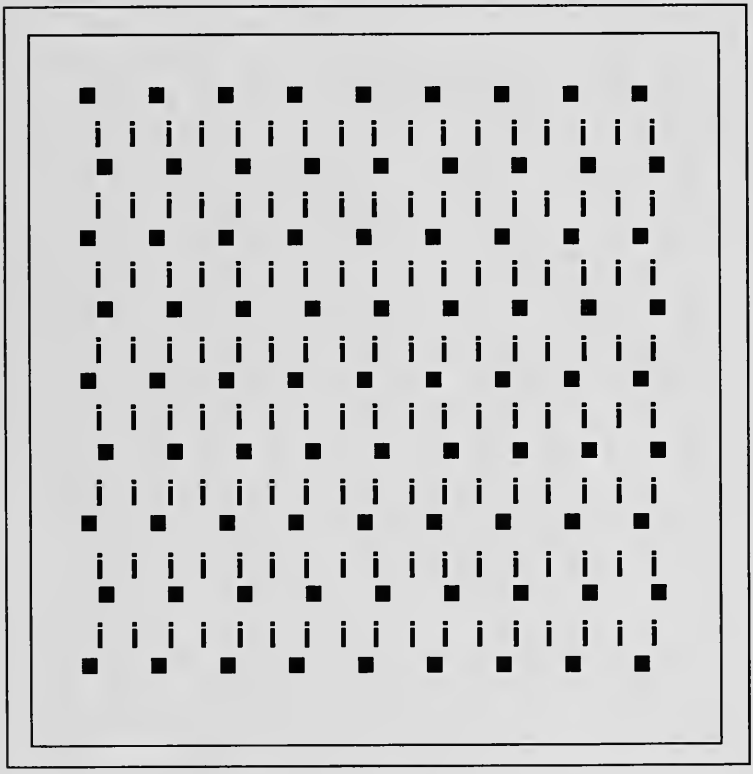
IASSIST

QUARTERLY

VOLUME 11 FALL/WINTER 1987 NUMBER 3/4



IASSIST QUARTERLY

VOLUME 11 NUMBER 3/4 FALL/WINTER1987

FEATURES

DEPARTMENTS

INTERNATIONAL ASSOCIATION FOR SOCIAL SCIENCE INFORMATION SERVICE AND TECHNOLOGY

ASSOCIATION INTERNATIONALE POUR LES SERVICES ET TECHNIQUES D'INFORMATION EN SCIENCES SOCIALES

Colonial control information available for Inuit family research

Disko Bay church and census records

The following three papers were presented at the IASSIST '87 conference in a plenary session entitled Research trends . The object of the session was to focus on the effect of new trends in research and data collection and the advance of technology on the use of data and data management techniques.

by Per Nielsen[2]
Danish Data Archives

[1]Presented at the International Association for Social Science Information Service and Technology (IASSIST) Conference held in Vancouver, British Columbia, Canada, May 19-22, 1987

[2]I am indebted to Kirsten Elisabeth Caning for the articles she sent me, and to Jens Ludvig Wagner at the DDA for assistance and support when I tried to get an overview of the project.

1. Historical Sources from Greenland

Indigenous people under colonial rule were not left in peace! This was not because of any sincere interest in the ethnic peculiarities of the people, but rather because bureaucratic registration also served the purpose of ensuring that the path of development was in accordance with the stipulations of the colonizing government.

In the case of the Eskimo population in Greenland, the Government was in Copenhagen, where different authorities requested very detailed information on personnel resources, production, buildings, etc. In most cases, the information was required in order to evaluate economic development; however, even health conditions, births and deaths, including reasons for death, were reported. And in realm of the church, baptisms, confirmations, weddings, and burials were reported to Copenhagen.

Prior to 1774, the "colonial initiative" was mainly a private one, consisting of tradesmen on one hand, missionaries on the other. In 1774, the main administrative tasks were centralized within Royal Greenland Trade, leaving the church to take care of its own affairs. From 1782 to 1950, Greenland was administratively divided into two Inspectorate Regions ,North and South, each with its own Inspector; the division line was located between Egedesminde and Holsteinborg. Needless to say, there were modifications introduced to this system during the last decades of the 19th century; and many administrative changes took place during the first half of the 20th century. However, a detailed description of these is beyond the scope of this short introduction.[3]

[3]The above outline is based on Kirsten Elisabeth Caning, "Personalhistoriske kilder i grønlandske arkiver." Personalhistorisk tidsskrift, 1979.

During the period preceding 1951, all church matters as well as most issues concerning education were governed by the same administrative unit in Copenhagen. It is from this clerical administration that most of the information dealt with in this article originates.

2. The Church Records

Church records were maintained in the Greenland parishes during the first half of the 18th century; most of them have disappeared, and a couple of preserved manuscripts are incomplete in that they include information on the top strata of the Greenland population only. From the second half of the 18th century, there are preserved church records from Upernavik, Godhavn and Egedesminde. The lists of those baptized (Christening Lists) represent the majority of entries in the early years. The salvation of the souls of the heathens was one of the major preoccupations of missionaries; consequently, the number of baptisms performed was an indication of the efficiency of the missionaries.

The church records from the beginning of the 19th century, held in the Danish archives, are more complete, especially those from many of the parishes in the North Greenland Inspectorate. The records from the South Greenland Inspectorate are not so complete, in part due to the loss of a lot of these documents during their transportation to Denmark aboard the ship "Hans Hedtoft" (January 1959).[4]

The church records which have been preserved consist of four separate types of lists representing four different types of events: Christening Lists, Confirmation Lists, Marriage Lists, and Burial Lists. Concommitantly, the data sets stored at the DDA (the "raw data") are separated, in their initial form, into 4 different files:

- DDA-0311: Population History of Greenland 1800–1930: Christening Lists
- DDA-0312: Population History of Greenland 1800–1930: Confirmation Lists
- DDA-0313: Population History of Greenland 1800–1930: Marriage Lists
- DDA-0314: Population History of Greenland 1800–1930: Burial Lists

3. The Census Records

The missionaries were also responsible for the registration of all people in their districts (including those who were not Christians, and who, consequently, never appeared in the Church Records). The censuses in Greenland contained the same information as those in Denmark: name, age, occupation, and position in household. In addition to these, the Greenland censuses contained information on race (Eskimo, mixed and European) as well as an indication of whether or not the person had been baptized.

At the National Archive in Copenhagen, there are census records from: 1834, 1840, 1845, 1850, 1860, 1870, 1901, and 1911. Those from the Umanak and Disko Bay Districts, for all censuses except the 1911 census (because of the 80-year access restriction to personal records at the National Archives), have been made computer-readable and are available as:

[4]In the "Danish Titanic" case, "Hans Hedtoft" collided with an iceberg on her maiden voyage to Greenland, January 30th, 1959. All 95 crew members and passengers died.

- DDA-0645: Population History of Greenland 1800–1930: Census Lists

3. Geographic Coverage and Time Period Included in the Datasets

In the initial definitive stages of the "Demographic History of Greenland, 1800–1930" research project, both geography and time periods had to be taken into account. The above outlined existence of nearly complete archival series of church records and census records in Archives situated in or near Copenhagen, has some bearing on the geographic regions and time periods to be covered.

Geographically, the North Greenland Inspectorate had the most complete records. Consequently, the rather isolated regions in northwest Greenland, the Disko Bay District and the Umanak District (isolated in terms of in- and out-migration from/to the surrounding regions, so that few individuals "disappear" via migration) were selected; from north to south (or rather, "around the bay") in the Umanak and Disko Bay Districts thus defined, Umanak, Godhavn, Ritenbenk, Jakobshavn, Christianshåb, and Egedesminde are the larger township areas included in the project.

With respect to time period covered, the availability of records suggests that the registration should begin around the turn of the 18th century, i.e. around the year 1800. For reasons of restricted access (and, of course, as a means of reducing the resources required for the data collection process), the registration was stopped at the beginning of the 20th century.

4. The Data Collection Process

Coding was done from the original sources, or copies thereof by the principal investigator, Kirsten Caning, together with a student aide (see figure 1).[5] The coding sheets were then sent out to have the information transformed to a computer-readable medium. The machine-readable data were deposited with the DDA, where Jens Ludvig Wagner (who was experienced in demographic data and family reconstitution methods from his work with several other historical projects) took over the transformation tasks as they were requested by Ms. Caning.[6] During this process, the following rectangular files were generated:

- DDA-0311:17,248 Baptism entries, each with up to 53 variables
- DDA-0312: 10,037 Confirmation entries, each with up to 62 variables
- DDA-0313: 4,720 Wedding entries, each with up to 90 variables
- DDA-0314: 13,217 Burial entries, each with up to 84 variables
- DDA-0645: 23,953 Census entries, each with up to 37 variables

These five files form the basic "raw data" of the project. However, after this initial collection of raw data, a tremendous effort was devoted to data correction and family reconstitution. Correction involved, for example, the deletion of double entries (e.g. the baptism of the same child in two parishes in the Church Records, or

[5] [figures and tables are collected together at the end of the article. Ed. note]

[6] See Jens Ludvig Wagner, "Datamaterialer med komplekse strukturer", in DDA-Nyt 25:64-70, 1983.

enumeration of the same person in two households in the Census Records). Those with experience in family reconstitution projects know the amount of preparatory work to be done before one can produce a centralized file. Therefore, I shall go into some detail concerning the methodology applied in this particular family reconstitution project. How did we derive, from the above listed five raw data files within which all the events were sequentially numbered, the so-called "centralized file" containg the best possible reconstitution of families?

5. Family Reconstitution: Manual and/or Automatic Procedures

Within the discipline of historical demography, there has been quite a long tradition of two "schools", one advocating automatic (i.e. computer based) family reconstitution procedures, the other maintaining that a lot of human thinking is necessary in order to get as close to the ideal of a complete reconstitution as possible. In the case of the data from Greenland, a mixed automatic and manual process was used.

With the raw data (or basic files) as the point of departure, two tasks had to be completed: that of identifying the events (e.g. a baptism, confirmation, wedding, burial, or membership in a certain household in a specified census enumeration) as attributes, or descriptors, characterizing the "central persons" (or "actors") when the event took place, and that of establishing "pointers" among all the central persons based on their family relationships.

Who are the Central Persons (CPs) in this project? - In the event of a baptism or a confirmation, the baptized/confirmed person and his/her parents are the CPs; in the event of a wedding, bride and groom as well as parents of both of the married persons are CPs; when the event is a burial, the buried person and his/her parents as well as a spouse and children are CPs. Finally, in census enumerations, each person listed in each census is a CP. By machine generation, the Basic Files were thus expanded into a "Theoretical Centralized File" with more than 200,000 "central person in one event"-combinations. We shall call these Central Person-Event records (CPE). A sequential number was allocated to each CPE.

The CPE-file was sorted on names; some auxilliary lists sorted on districts and other criteria were produced by Jens Wagner whenever Kirsten Caning needed and requested such lists during the work. The sorted CPE-file(s), written out as long paper listings, formed the basis for the next major state: the process of manually inserting Personal Identification Numbers (PIN), (see fiqure 2).

Even though the sorting based on names did help a lot, the task of numbering more than 20,000 individual persons (as it turned out later) on more than 200,000 CPE-records was a heavy consumer of both time and human memory! And now we may return to the question whether the machine or the manual reconstitution is "better". It goes without saying that the choice of method depends heavily on the nature and quality of the raw data; so we shall talk about advantages and disadvantages, (see fiqure 3).

Automatic family reconstitution has the advantage of being relatively quick and being well documented. However, in many cases a lot of persons are left "split" in two or more persons. For example, figures 1 and 2 (1901-census) compared to the extensive figure 3 (an excerpt from the CPE-file after all CPs have been numbered) clearly demonstrates a number of problems with machine reconstitution: Nikolaj Jens Andreas Lange was found under many different names, partly due

to the fact that the first names may be written in any order (some of them may be missing), partly because of different spellings (some of which are not caught in the normalization process, e.g. Nik(olaj) vs. Nic(olaj)). Figures 2 and 3 show that Kirsten Caning preferred to work with "normalized names" in order to minimize the spelling problem. But even with that precaution, the decision was made to perform the reconstitution manually.

The reconstitution process was carried out by entering PIN-numbers on all relevant locations in the CPE-file. During that process, a number of errors were detected and corrected - this is probably one of the main advantages of manual reconstitution. After some iterations of PIN-numbering, the theoretical CPE-file was reduced from more than 200,000 records to approximately 131,000 CPE-records.

6. The Documentation Problem in Manual Family Reconstitution

Who "disappeared" from the theoretical CPE-file? Approximately 70,000 entries vanished from the "theoretical" to the "reduced" version of the CPE-file. The records of type "own baptism" were reduced from 17248 to 16312; "own confirmations" were reduced from 10037 to 9169; "own marriage" from 9442 to 8388; and "own burial" from 13217 to 11678. Because some census registrations had been erroneously omitted from the first data entry process, the number of census registrations increased from the original 23953 to 24076. (The first number is the size of file given in section 4 for Raw Data Files).

Theoretically, one might expect that every person involved in a baptism, confirmation, marriage, or burial would have two parents. However, these CPs have, in many instances, not been identified in the source records. Therefore the number of CPE-records in which parents were registered in each of the four events was reduced from the "theoretical" numbers 34496, 20074, 18884 and 26434, to 31338, 13538, 4902 and 9181 respectively. Predictably, it was in the events in which the "main actors" were adults (marriages and burials) that the information on parents was most commonly missing.

There is a major difference between the theoretical expectation that there was a spouse for each person buried (i.e. 13217 spouses) to the actual 2328 cases in which a spouse was actually identified. This is due, to a great extent, to the fact that many burials were of children (high infant mortality rate), and also in part to the death of unmarried adults; also some of the missing spouses were due to lack of registration. Finally, theoretically, it was expected that there would be information on at least one child for each buried person (i.e. an expectancy of 13217 persons); however, it turned out that this information was available in the source records only in 56 cases.

Does this mean, now, that we have lost the information in those of the CPE-records that were not merely theoretical, but in which the information was not given in the source records? It does not!

If a person was buried who was married at the time s/he died without an indication thereof in the burial record, we have data about the marriage from a different source. Similarly, if one or both parents of a baptized person were not mentioned in the Baptism File, we still may know of them from other sources, e.g. from a relevant census. This leads us on to the "record linkage", which in this project was done in an *ad hoc* data system.

7. Finalization of the Family Reconstruction in a Data Base Environment

During 1984 and 1985, several MLD/ASTRID[7] data bases were designed and tested on subsets of the data from the Greenland Project. Even though a number of these designs were acceptable from the substantive point of view, they had to be rejected because they would consume too much computer power when loaded with the full data base. [8]

With further computations and tests, a data base design was finalized by the end of 1984[9] and implemented with all the data from the project. The basic design of the data base is that one part contains all PERSONs and another part contains all the EVENTs. Each person is identified by the PIN-number preceded by M (males) or K (females). There is no need for a lot of pointers because the searching is based on key fields in PERSONs and EVENTs respectively.

I shall not go into the technical details of the data base in this paper. However, I shall demonstrate the output from the data base using the two heads of household that we have followed in figures 1-3 above, (refer to figure 4).

As can be seen from figure 4, not much has been done to present the output in an easily understandable way; thus far, only people actively involved in the project have used the system, and they know what the encryptions mean. If the data base system were to be used by others outside the project, more text would have to clarify the output.

Using this data base, the principal investigator is presently finalizing the family reconstitution task. Also, special purpose functions have been designed for demonstration purposes; for example, it is quite easy to establish pointers to allows a user to move up and down along genealogical lines, drawing the complete pedigree of the person under analysis.

It should be noted here that several features that are specific to the Inuit culture are reflected in the data. People may be referred to their biological family, to an extended family or household, or to the particular house in which they lived. It was quite usual for the Eskimo to live in so called longhouses.Several households lived in the same longhouse during the winter (when censuses were taken); during the summer, they might move away from the house, living in tents or similar dwellings. The following year, the family might move to a different longhouse at the same or at a different location; the houses were not owned by anyone - or, rather, they were owned by whoever happened to be occupying them. This tradition of living together in longhouses included a social security aspect; the data show that this tradition was slowly abandoned in favour of nuclear family houses during the 19th century.[10]

With the huge number of cross-identifications in the data, it is possible to follow individuals or families (biological or extended) for generations. Alternately, it is possible to examine a single location, and to describe how life changed in that particular location over the

[7]MLD/ASTRID is a data base language and system, defined by Jørgen Grosbøl at DDA, described in his manual: MLD Data Bases and the ASTRID Language. Odense: DDA 1981.

[8]The basic design was described in Karsten Boye Rasmussen, "MLD/ASTRID database for Grønlands befolkningshistorie", Working Paper A460-KB. DDA 1984-07-31.

[9]Jens Wagner, "MLD/ASTRID database for Grønlands befolkningshihistorie - design uden referencer", Working Paper A460-JW, DDA, 10. December 1984.

[10]Described in Kirsten Caning, "Fra bofællesskaber til kærnefamilie", Beretning fra Carlsberfondet, København 1986.

years. The latter strategy was applied by the principal investigator in an article on Sermermiut – a small place with only two houses with 20 and 12 inhabitants respectively (3 households per house).[11]

8. Access to the Datasets for Secondary Analysis

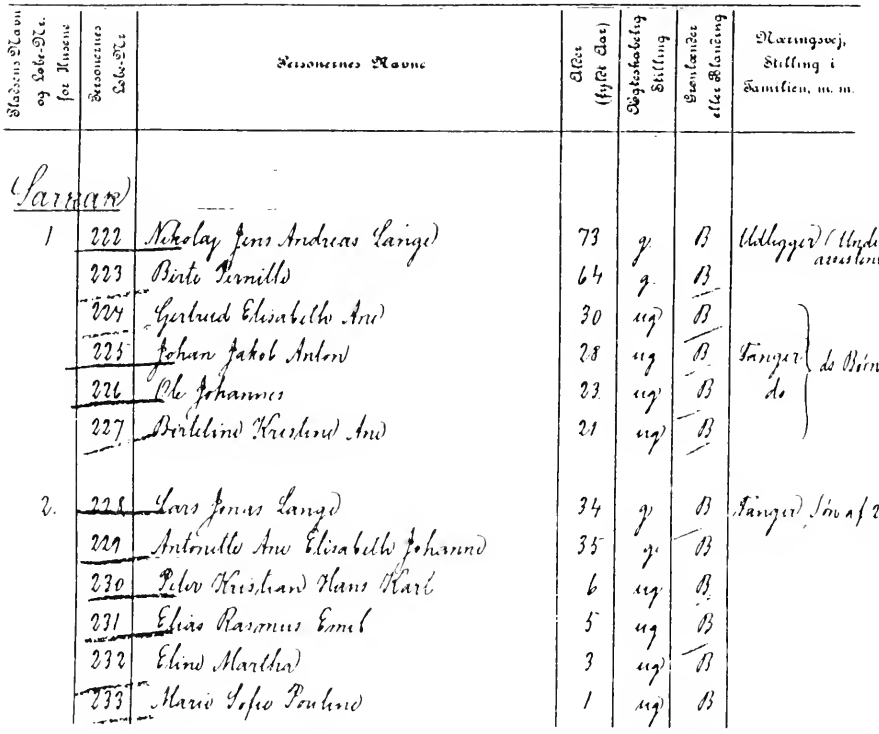
Despite the fact that the primary investigator has not yet finalized the troublesome reconstituition process, it will be possible for secondary analysts to have access to the data, with the consent of Kirsten Caning. All requests should be directed to the DDA.

Needless to say, it is not quite as simple to address a user requesting this type of data as it is to send out a survey file. The user must define *a priori* the subsets or the formats s/he can handle. Further, it is the nature of historical-demographic data of the type we have been discussing that their quality improves over the years; during analysis, new findings concerning data relationships can be added. Such changes are now reflected in the data base version of the data, but not in the Raw Data Files mentioned in section 4.

Consequently, disseminable versions of the data should be extracted from the data base according to the specifications of the individual user. With the data stored in a data base management system, it is possible to generate rectangular files that can be analyzed with standard statistical software. However, to get the full personal history description capabilities, a data base management system environment would be needed by the user.□

[11]Tinna Møbjerg and Kirsten Caning, "Sermermiut in the Middle of the Nineteenth Century", Arctic Anthropology vol. 23(1-2):177-198, 1986.

figure 1



Husets Navn og Løb-Nr. for Husene	Personernes Løb-Nr.	Personernes Navne	Alder (fyldt Aar)	Ægteskabelig Stilling	Sædvanlig eller Midlertidig	Næringsvej, Stilling i Familien, m. m.
(Sarraaq)						
1	222	Nikolaj Jens Andreas Lange	73	g	B	Udlagger (Undersøgelsesrejsende)
	223	Birte Pernille	64	g	B	
	224	Gertrud Elisabeth Ane	30	ug	B	
	225	Johan Jakob Anton	28	ug	B	Fanger hos Bøtø
	226	Ole Johannes	23	ug	B	do
	227	Berthelina Kirstine Ane	21	ug	B	
2.	228	Lars Jonas Lange	34	g	B	Fanger Søn af 222
	229	Antonette Ane Elisabeth Johanne	35	g	B	
	230	Peter Kristian Hans Karl	6	ug	B	
	231	Elias Rasmus Emil	5	ug	B	
	232	Elina Martha	3	ug	B	
	233	Marie Sofie Pauline	1	ug	B	

Figure 1: Xerox copy of one of the source materials, viz. the Census List from October 1901, with information from Saraqaq in Ritenbek. The upper household is that of Nikolaj Jens Andreas Lange, living with his wife and 4 unmarried children aged between 21 and 30. The lower household is an older son, Lars Jonas Lange, with his wife and 4 children.

The figure is reproduced from an illustration in Kirsten Elisabeth Caning, "Om den grønlandske befolknings historie", printed in Forskning i Grønland 1/82, p. 5.

figure 2

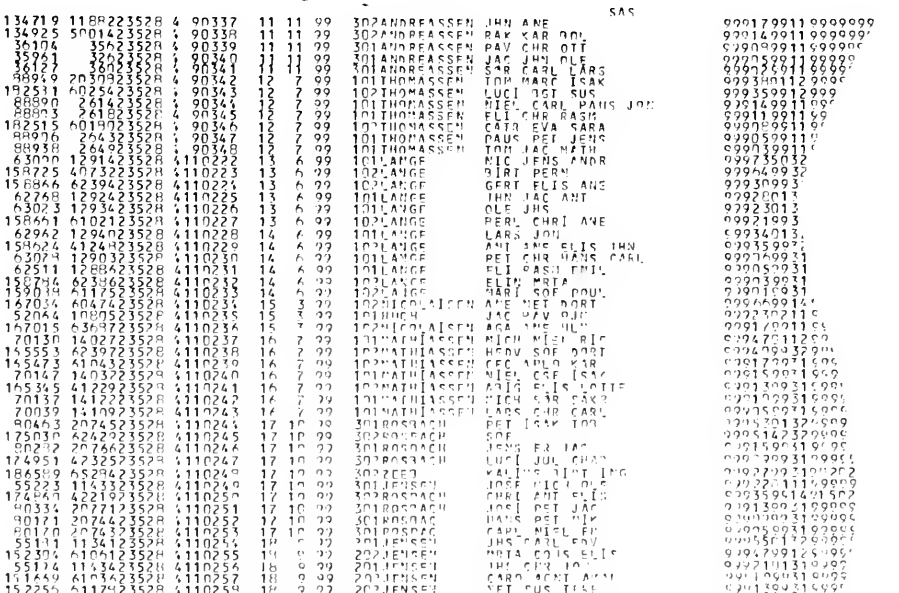


Figure 2: The same persons as those listed in figure 1, now listed from the computer. To the left of the "normalized" names field, lots of numbers of identification appear, i.a. the PIN-number appearing in columns 7-12. Our head of household Nikolaj Jens Andreas Lange has PIN-code no. 12914 - in the Ritenbenk District (code 4 in col. 19), the 1901-census (code 28 in cols. 16-17) from which the information was taken. He has that number in all other datasets as well, which we shall see later.

figure 3

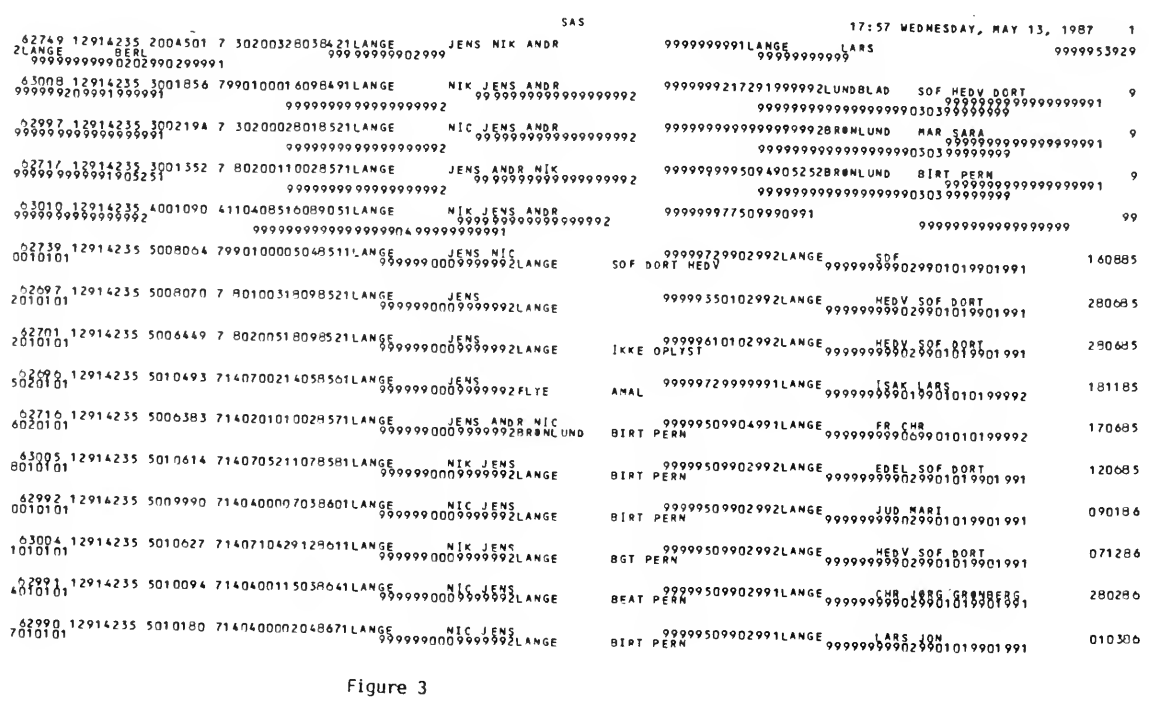


Figure 3

Figure 3 ...cont

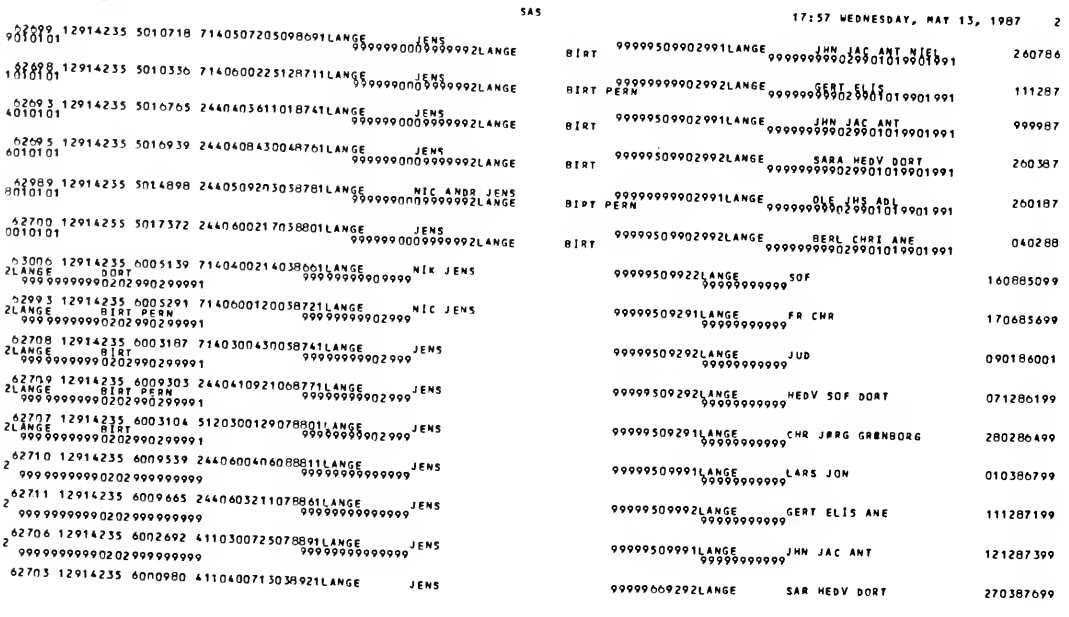


Figure 3

figure 3 ...cont

```
2                                         99999999999999                SAS                          99999999999 17:57 WEDNESDAY, MAY 13, 1987     3
  99999999990202999999999
 62996 12914235 6000875 41104009130389 21LANGE       NIC JENS ANDR           99999669291LANGE          OLE JHS ADL                 260187899
2LANGE       BIRT PERN                    99999999902999                                    99999999999
  9999999990202990299991
 62704 12914235 6000995 41104023150889A1LANGE        JENS                    99999509992LANGE          BERL CIRK ANE               040288099
2                                         99999999999999                                    99999999999
  99999999990202999999999
 62995 12914235 8007027 71407051291086 01LANGE       NIC JENS                999999999509299992LANGE       SOF DORT                  12
06858999999999 12LANGE           BIRT PERN      999999999999299991                                   999999999999999999
                                   9999999999999999904990404 99991
 62712 1291A235 8007214 71405120220787 11LANGE       JENS                    999999999509299991LANGE       JHN JAC ANT NIEL          26
07869999999999 12LANGE           BIRT           9999999999999299992                                  999999999999999999
                                   9999999999999999904990404 99991
 62714 12914235 9003467 41105067010790 91LANGE       JENS                    999999999509699992LANGE       BIRT PERN                 99
99999729994 90991                               9999999999999999992                                  999999999999999999
                                   99999999999999999904 9999990 4991
 63011 1291423521 7 6 197   20   9   2   301LANGE    NIK JENS ANDR           82507993129999999 0 0 0 1 2   49  5 3 0 0  3   508 3  5

 62998 1291423522 7 6  951000100100 10001LANGE       NIC JENS ANDR           82815993129999999 1010101010  41  4 2 0 0  5   517 2  4

 62999 1291423523 7 6  811000100100 10001LANGE       NIC JENS ANDR           82818993120500759 1010101010  40  7 3 2 0 12   522 5  1

 62741 1291423524 7 8  201000100100 10001LANGE       JENS NIC                99922725229999999 1010101010  70  2 1 0 0  1   535 1  1

 62742 1291423525 7 8  21   25   4  99   501LANGE    JENS NIC                99928723429999999 0 0 0 3 010000100101010100  561 6  1

 63007 1291423526 714 255100010010 010001LANGE       NIK JENS                99932505229999999 10101010100001001010 10100  659 3  1

 62743 1291423527 714 515   43  12  99   301LANGE    JENS NIC AND            99942505229999999 0 0 0 2 110000100101010100  858 4  1

 63000 1291423528 411 222   13   6  99   101LANGE    NIC JENS ANDR           99973505299999991 0 0 0 0 110000100101010100  800 3  1

 62958 12940235 1010180 7140400002048671LANGE       LARS JON                 0103867010101                                        9999990
009999991LANGE      NIC JENS               99999509902992LANGE       BIRT PERN           9999999990299010199 01991
```

Figure 3

figure 3 ...cont

```
 62959 12940235 2009339 2440600406088811LANGE        LARS JON          SAS           0103867991LANGE        17:57 WEDNESDAY, MAY 13, 1987     4
2                                         9999999999999                                    9999999999 JENS                             9999950999
  99999999990202999999999

 62960 12940235 3000407 41104020050889 41LANGE       LARS JON                        99999992701919040062CHRISTENSENANT ANE ELIS JHN KAR     9
99999927999 1904241                                      9999999999999999992            99999999999999991
                                   9999999999999999992                                     9999999999999 99999030399999999

 62942 12940235 5001590 411040392 30689 51LANGE      LARS JON                   9999901990299 1LANGE        PET CHR HANS CARL          280589
5010101                                       999999000999999 2LANGE      ANT ANE ELIS          9999999990299010199 01991

 62943 12940235 5001599 411040480 701 8971LANGE      LARS JON                   99999019902991LANGE         ELI RASM EMIL              091289
6010101                                       999999000999999 2LANGE      ANT ANE ELIS          9 99999999029901019901991

 62944 12940235 5002055 411040561 3118981LANGE       LARS JON                   99999019902992LANGE         SAB HLN MRTA               181089
8010101                                       999999000999999 2LANGE      ANT ANE ELIS JHN KAR  99999999902990101 9901991

 62945 12940235 5002071 41104072111119001LANGE       LARS JON                   99999019902992LANGE         MARI SOF PAUL              251090
0010101                                       999999000999999 2LANGE      ANT ANE ELIS JHN      99999999902990101 9901991

 62946 12940235 5001670 41104119170990 51LANGE       LARS JON                   99999019902991LANGE         JENS NIK ANDR              200890
5010101                                       999999000999999 2LANGE      ANT ANE ELIS JHN      99999999902990101 9901991

 62863 12940235 6003549 714041090108909 1LANGE       LARS                       99999509291LANGE      PET CHR HANS CARL             280589599
2LANGE          ANE                     99999999902999                                  99999999999
  99999999990202990299991

 62948 12940235 6001956 714060072506911 1LANGE       LARS JON                   99999509291LANGE      ELI RASM EMIL                 091289599
2LANGE          ANT ELIS JHN            99999999902999                                  99999999999
  99999999990202990299991

 62949 12940235 6002115 714060181708913 1LANGE       LARS JON                   99999509292LANGE      SAB HLN MRTA                  181089799
2LANGE          ANT ELIS JHN            99999999902999                                  99999999999
  99999999990202990299991

 62950 12940235 6002127 714060302502916 1LANGE       LARS JON                   99999509292LANGE      MARI SOF PAUL                 251090099
2LANGE          ANT ELIS JHN ANE        99999999902999                                  99999999999
  99999999990202990299991

 62947 12940235 6002007 714060562803920 1LANGE       LARS JON                   99999509291LANGE      JENS NIK ANDR                 200890599
2LANGE          ANT ELIS JHN ANE        99999999902999                                  99999999999
  99999999990202990299991

 62951 12940235 7000881 7 80602814029161LANGE        LARS JON                   999999999019699999 1LANGE       JHN PET CARL SAKR       1
4048899901919070 12JENSEN      MGDL SARA FR          08018989999919070 12LANGE         CHRI                99999999999 4999991JENS
EN        SIGV DLE VILS CASP          99999999901969999 2JENSEN     JHN KAR MRTA        9999999999999 49999030303030303

 62954 12940235 7000740 51206045090491 61LANGE       LARS JON                   99999999950929052 51LANGE       PET CHR HANS CARL       2
80589599539190501 20LRIK       BATZ REG MARG         2812890996191905012CHRISTENSENANT ELIS JHN ANE    999999999999290525 10LRI
K         CHRI FR CARL                99999999950949051120LRIK        BIB              9999999999999 690511030303030303

 62952 12940235 7000898 8 50604521049201LANGE        LARS JON                   99999999950929070 31LANGE       ELI RASM EMI            0
912896990191907032LYNGE        KAR JAC PET AUGN      02108999999919070 82LANGE     ANE ANT ELIS        9999999999999290705 1LYNG
E         AUGN                        99999999983969070 82LYNGE       JUL KAR BIRT     9999999999999 490708030303030303
```

Figure 3

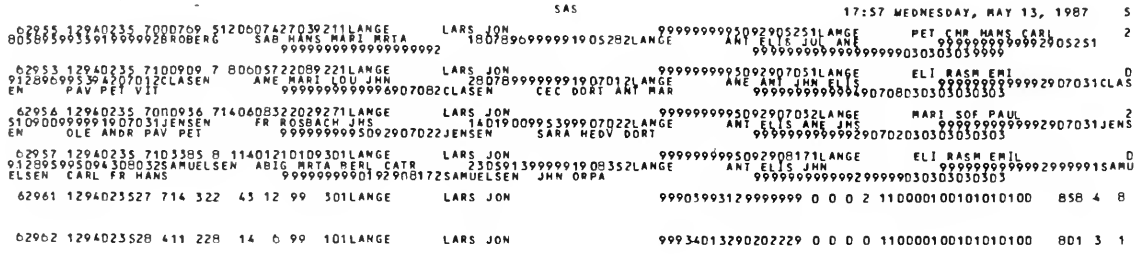


Figure 3 (pp. 10–14 incl.): Excerpt from the CPE-file after the PIN-codes have been entered. The excerpt covers the two heads of household that were listed in figures 1 and 2, i.e. PIN-nos. 12914, Nikolaj Jens Andreas Lange, and 12940, Lars Jonas Lange. Here is a short introduction to the two persons:

12914 Nikolaj Jens Andreas Lange: The first block of 3 lines of this figure concerns his confirmation (code 2 in col. 17 of the first line); strangely enough, he seems never to have been baptized. The following 3 blocks (code 3 in col. 17) concern his 3 consecutive marriages, with Sof(ie) Hedv(ig) Dort(e) Lundblad, Mar(ia) Sara Brønlund, and Birt(e) Pern(ille) Brønluns – to whom he has earlier been shown to have been married in 1901.

After the three marriage records follows his burial registration (4 in col. 17 of the first line). The following blocks (code 5 in col. 17) contain information on the baptisms of 16 children, and confirmation records on 11 of the children (code 6 in col. 17). There have been burials of two of his children (code 8 in col. 17), and burial of one spouse (code 9 in col. 17). Finally, Nikolaj Jens Andreas Lange is present in every one of the 8 census enumerations, reflected in blocks with codes 21–28 in cols. 16–17.

12940 Lars Jonas Lange: Following the same coding patterns, Lars Jonas Lange is seen to have been baptized, confirmed and married (codes 1, 2, and 3 in col. 17). 5 children have been baptized and 5 confirmed (codes 5 and 6 in col. 17); as many as 7 events are registered with marriages of his children (code 7 in col. 17), and Lars Jonas Lange has appeared in 2 censuses (codes 27–28 in cols. 16–17).

figure 4

```
*PERSON: M012914
BARN        K062355 M012938
BARN       =M012927 M012891 K062429 K062398 K040998 M012878 M012940 K062394
           M012906 K062384 K062397 K061204 K061021 M012934 M012924
GIFT        K040732 K040712 K062382
STAMDATA    1825 JENS NIK ANDR LANGE 3 9 1905 816
TRANSDATA   B010901905081601500 B034671909070102506 B063741860102920502
            B072141871072210502 D063831857021010504 D080641851040520725
            D080701852091820352 D099901866030720502 D100941864031510502
            D101801867040210502 D103361871122520002 D104931856051410720
            D106141858071120502 D106271861122920502 D107181869090510502
            D168351878031031002 D065801874010410502 D169391876043020502
            D173721880031720502 K008751892031310662 K009801892031320662
            K009951894081520500 K026921889072510500 K031041880072910502
            K031671874053020502 K045011842032801000 K051391866031420500
            K052011872032010502 K093031877062120502 K093391881080610500
            K096651886071120500 S700401845000012001 S700411840000003001
            S700491834000003001 S700701850000001722 T408091901000001502
            T705081834000003001 T705171840000003001 T705221845000001001
            T705351850000001722 T705631855000001724 T706591860000001502
            T708581870000001722 V013521857021001504 V018561849091601721
            V021941852012801000
*PERSON: M012940
BARN        K040732 M012914
BARN       =M012903 M012886 M012916 K062386 K061175
GIFT        K041248 K061998
STAMDATA    1867 3 1LARS JON LANGE 3 11
TRANSDATA   D015901895062310012 D015991897010710012 D016701905091710012
            D020551898111320012 D020711900111120012 D101801867040201000
            K019561911062510502 K020071920032810502 K021151913081720502
            K021271916022520502 K035491909080110502 K093391881080601000
            T408111901000001012 T708581870000003001 V004071894080501011
            V007401916040910502 V007691921032710502 V008811916021410016
            V008981920042110502 V009091922082210502 V009361927022220502
            V033851930101010502
```

Figure 4: The information in the PERSON archive on the same two heads of household as reflected in earlier figures. The person is identified by the PIN-code preceded by M (males) or K (females); the second line gives the parents, third (and forth if necessary) the children. Under the heading GIFT (= married) the spouses are identified. Finally, under the heading TRANSDATA there are references to all events in which the individual was considered a CP. The initial letters identifying events are: B for burials, D for baptisms, K for confirmations, S for censuses with the extended family as the unit, T for censuses with a biological family as the unit, and V for the weddings.

Canadian social trends:
a new Statistics Canada social statistics publication

by D. Craig McKie, Ph.D.[1]
Chief, Social Reporting & Editor, Canadian Social Trends
Housing, Family and Social Statistics Division
Statistics Canada

In the last ten years at Statistics Canada, the governmental collector and keeper of the nation's storehouse of numerical information about itself, momentous changes have occurred in the ways data are collected, processed, analyzed and published. It is not recognizably the same organization I first joined in 1977. Since that time, the widespread importation of new methods and new technology have in turn been reflected in organizational changes. Accompanying these changes have been pressures to reduce costs and overall staff levels and to recover a larger part of the costs of operation from clients.

While the technical aspects of this change are perhaps better known (and we could refer to random digit dialing-based interviewing techniques, a microcomputer on every other desk, and very sophisticated analytic software and graphics hardware to match it), changes on the publishing side have also been considerable. But while the necessity for modernizing the dissemination of data has been well recognized (by, for instance the widespread acceptance and use of electronic bulletin board type data services and the Bureau's own service CANSIM, and by sales of data on floppy disks, user tapes, and other similar services), publishing has not received quite the same amount of attention. The Bureau, for instance, still publishes many strange and exotic publications such as: the Cereals and Oilseeds Review, Gas Utilities, The Dairy Review, and last but not least, Sawmills East of the Rockies.

However, the number of users of whatever origin who are ready to plow through table upon table to find a single figure – or to construct their own annual time series from 25 years of monthly pamphlets is limited. Further, many users have neither the time, the training, nor the inclination to do such work.

Those who decline to engage in such activities are legion. Some telephone me; undoubtedly some may telephone some of you. Indeed, at

[1]Presented at the International Association for Social Science Information Service and Technology (IASSIST) Conference held in Vancouver, British Columbia, Canada on May 19–22, 1987

the recent meetings of the Population Association of America in Chicago, one eminent demographer, who shall remain unnamed, stated to one of my colleagues that it was easier to obtain Zambian data than to get it from Statistics Canada.

Such reactions are not uncommon. Foreign nationals, or the inexperienced, simply do not know where to begin. Other domestic users, although they know how to begin, lack the iron will necessary to persist in their quest. They tend to fall into several broad categories of persons. Three prominent types are:

1. those who are interested in a subject but lack the knowledge of how to go about finding suitable data and who are intimidated by the traditional publication of official statistics [massive cryptic tables with lots of footnotes, runic symbols, and qualifications which seem meant to undermine any simple interpretation of the contents];

2. those who know what they want and where to find it but cannot afford the time to assemble the data nor interpret it [there are few who do not shudder at the task of sorting through 52 pages of tables, the numbering of which may change from year to year, over enough years to assemble a graph which can indicate the direction of change, perhaps standardized to some rate per population];

3. those who don't want data at all but rather a prose summary of a data set to use as evidence.

I think it was our belief when we started to think about the form of what turned out to be CST, in 1984, that most Canadians, to the extent that they seek to use Statistics Canada material, fall into one of these three categories or subsets of them. To reach such people with useful information, a new format would be necessary. Our subsequent activities in search of such a format have been based on the premise that there is a significant demand from a wide range of governmental and non-governmental agencies for comprehensive and systematic information which identifies, describes, and analyzes social trends and conditions in Canada. This demand is expressed in a number of ways including requests to Statistics Canada for assistance, information, and ideas. These requests come from government agencies, members of parliament, community groups, and the press. Canadian Social Trends is one response to these requests.

There is much potential for information which illuminates the social conditions which are the objects of policies, programmes, and public concern. After some trial and error, which involved the preparation of a similar annual format (it was discarded because of the inevitable loss of timeliness in an annual publishing cycle), the final decision was to go with a quarterly format which at least superficially looks like a news magazine – a format familiar to almost everyone in western countries and one users expectations of which are in a sense built-in. One expects succinct, highly graphical treatment of issues pertinent to the reader (they are not arcane) interspersed with advertising (information).

The start of publication of Canadian Social Trends is one notable aspect of the latter type of change.

The New Publication

In terms of the number of its publications (in excess of 600) Statistics Canada could well be the largest publisher, public or private, in Canada. However, its publications have by tradition been dry and heavily given to technical

detail. In contrast, Canadian Social Trends is an entirely new type of publishing endeavour for Statistics Canada.

The purpose of CST is to report on changing social conditions in Canada in order to provide government and corporate policy-makers and planners, marketers, and others with a basis for decision-making. It provides recent and historical evidence for the direction and magnitude of change of important social trends. Its intent is to describe and make evident a trend's inter-relation with policy concerns, and promote an understanding of social conditions (and economic developments where a direct social outcome is established).

The publication has five distinct objectives:

1. to provide decision-makers in government, industry, and the social service sector with an awareness of social and socio-economic trends and information relevant to their environment in an appealing format. Improving accessibility to Statistics Canada information and information products is the prime objective;

2. to provide users with a single, authoritative source for social trend description and analysis in an easily accessible and readable form. The intention is to integrate sets of data and make the connections between isolated findings;

3. to provice committed information users with information about Statistics Canada's data holdings, information products, and services;

4. to provide Statistics Canada with a flagship with which to raise public awareness of its corporate character, value, and contributions to the public interest. The publication supports the data collection process by showing a public return of information to respondents – this return helps justify impositions on respondents' time;

5. to provide a vehicle with which to publicize the existence of Statistics Canada information and data products and statistical services to get the greatest possible use from a considerable investment of public funds in data collection while at the same time earning revenues to help defray the costs.

Audiences foreseen or already realized for this publication are: policy-makers and planners, both governmental and corporate, social service agencies, educators engaged in teaching social or Canadian studies at the secondary or post-secondary level, libraries, the media, international information repositories, and market research concerns.

Subject matter is selected according to a loose formula. In each issue there should ideally be found one article on each of the following general topics: demography, labour force, and income. In addition, a rough rotation of issue selection exists with respect to institutional areas (e.g. culture, health, education, and justice); and with respect to target groups (i.e. women, the elderly, youth, native peoples, and the disabled).

New Presentation Styles

It is difficult in words to describe style. Therefore I will show you some before and after shots of the same data in the old and the new format. *(see examples collected together at the end of this article. Ed. note).*

Now entering its second year of publication, CST is intended to be the authoritative voice for Canadian social statistics as they are collected and disseminated by Statistics Canada. The quarterly publication attempts to place social statistical reporting on a much more accessible basis than was previously the case. It

carries analytical articles of varying lengths on important topics of social statistical reporting (such as the decline in fertility and the increase in lone parent families). Each edition of the report is approximately 44 magazine format pages long and features text written for the educated layman. With 25 – 30 graphics per issue (photographs and illustrative drawings), it is primarily intended to be used as the best source of Canadian social trends material for government policy makers, planners, and private sector marketing specialists in addition to students. There are no private sector publications which compete with CST. As a result, it has quickly become the obvious source for Canadian social statistical data and carries new analyses and data reporting in addition to publishing summary articles on previously released data. In addition, each edition carries a social indicators page and notices of the release of new Statistics Canada publications which might be of interest to readers.

What is really novel about this publication is that it is a clear departure from the usual 'multi-table and short descriptive text' format traditionally associated with governmental statistical publications. Early acceptance of the format has been much greater than initially expected: Canadian Social Trends already has the largest paid subscription of any Statistics Canada periodical publication and indications are that interest outside the country is high, perhaps reflecting the fact that for a foreign reader unfamiliar with Canadian social statistics, this publication is in most senses self-explanatory, carrying with the articles the definitions and data qualifications necessary for reuse without further reference to sources. A three hour audio-visual presentation is also available to clients for a fee when resources are available.

Underlying the publication is a time series database containing some thousands of series. At some point in the future, clients will be able to purchase collections of these series, on diskette in Lotus format, thus greatly shortening their data collection and processing times and reducing the cost of data acquisition accordingly.

The publication is put together by a full time staff of six people, including myself, with the assistance of a professional design firm and students from Canadian university co-operative programmes. At any given time, we have from two to four university students on work term assignments. They receive course credit for working for the project. In addition, we employ some contract writers and editors, although the bulk of the editorial content is generated from within Statistics Canada.

The publication is issued in both official languages. Page design is by a contract design firm.

Examples

I will use the remaining time available to me to show you some concrete examples of what we mean by social trends. *(see examples collected together at the end of this article. Ed. note).* They may be thought of as significant changes of direction in the measures of significant social institutional variables.□

BEFORE

– 104 –

TABLE 6 Extra sentence imposed on Convicted Persons by Nature of Indictable Offence
Total and Provinces* – Continued

TABLEAU 6 Peine additionnelle imposée aux personnes condamnées selon la nature de l'acte criminel
Total et provinces* – Suite

Indictable offence / Acte criminel	Extra sentence – Peine additionnelle							
	Total extra sentence / Total peine additionnelle	Leave / Sursis	Fine / Amende	Restitution and fine / Restitution et amende	Good and bound and to keep the peace and fine / Prison et engagement de garder la paix et amende	Prohibition and fine / Interdiction et amende	Other prohibitions / Autres interdictions	Other penalties / Autres peines
	Cases – Causes							
Criminal negligence in operation of motor vehicle – Negligence criminelle dans la conduite de véhicule à moteur	2	–	–	–	–	–	2	–
Impaired driving – Conduite dangereuse	18	–	–	2	–	–	17	–
Dangerous operation of vessel – Conduite dangereuse d'un navire	1	–	–	–	–	–	–	1
Use of firearm in offence – Usage d'une arme à feu lors de la perpétration d'une infraction	7	–	–	2	–	5	–	–
Use of explosive or noxious substance – Usage d'une matière explosive ou délétère	1	–	–	–	–	–	1	–
Arson – Incendie criminel	9	–	–	5	–	4	–	–
Property damage – Dommages matériels	1	–	–	–	–	–	1	–
Arson: disregard for human life – Incendie criminel: mépris pour la vie humaine	1	–	–	–	–	–	1	–
Against the person – Contre la personne	112	–	–	31	2	63	13	3
Breaking and entering & place – Introduction par effraction dans un endroit	69	–	–	26	3	11	–	–
Against property with violence – Contre la propriété avec violence	69	–	–	26	3	11	–	–
Theft – Vol	28	–	–	21	1	7	–	–
Fraud and forgery – Fraude et contrefaçon	30	–	–	23	–	7	–	–
Having in possession – Avoir en sa possession	11	–	–	8	–	24	–	–
Theft – Vol	251	–	–	186	3	52	–	–
Theft of motor vehicle – Prendre frauduleusement des pneus	2	–	–	–	–	2	–	–
Against property without violence – Contre la propriété sans violence	626	–	–	318	6	112	–	–
Other interference with property – Autres ingérences concernant des biens	23	–	–	16	–	9	–	–
Other than offences against property – Autres infractions concernant des biens	23	–	–	16	–	9	–	–
Support and uttering – Faux et usage de faux	24	–	–	17	–	9	–	–
Offences relating to currency – Infractions relatives à la monnaie	4	–	–	4	–	–	–	–
Against the property rights in currency – Actes portant atteinte à la monnaie	34	–	–	21	–	9	–	–
Attempt to commit and accessories – Tentative et complicité	10	–	–	5	–	5	–	–
Bodily harm, firearm – Lésions corporelles avec usage d'une arme à feu	2	–	–	1	–	1	–	–
Causing bodily harm and uttering – Infliction de lésions corporelles et usage de faux	2	–	–	2	–	–	–	–
Drug offences – Infractions relatives aux stupéfiants								
Unlawfully assisting to drive impaired – Conduite pendant que les facultés de conducteur sont affaiblies	3	–	–	–	–	–	3	–
Driving while disqualified – Conduite pendant l'interdiction	2	–	–	2	–	–	–	–
Driving while impaired – Conduite avec facultés affaiblies	78	–	–	–	–	–	78	–
Failing to submit sample of accident – Omission d'obtempérer sur le lieu d'un accident	6	–	–	–	–	–	6	–
Offences relating to corrupting morals – Actes tendant à corrompre les mœurs	2	–	–	2	–	–	–	–
Offensive weapons – Armes offensives	11	–	–	5	–	6	–	–
Perjury and false statements – Parjure et fausses déclarations	1	–	–	1	–	–	–	–
Prison breach, escape and rescue – Bris de prison, évasion et délivrance illégale	3	–	–	3	–	–	–	–
Public mischief – Méfait public	9	–	–	4	–	–	5	–
Other – Autres	118	–	–	82	–	30	6	–
Criminal Code – Code criminel	751	–	–	433	10	304	103	1
Food and Drugs Act – Loi des aliments et drogues	10	–	–	4	–	6	–	–
Narcotic Control Act – Loi sur les stupéfiants	30	–	–	21	–	9	–	–
Federal statutes – Lois fédérales	42	–	–	25	–	17	–	–
Grand – Total – Général	793	–	–	459	10	221	103	1

* Excludes Quebec and Alberta – Le Québec et l'Alberta non compris

– 105 –

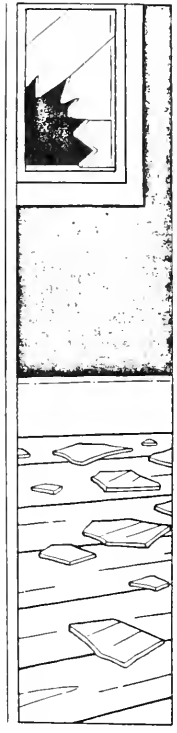
TABLE 6 Extra sentence imposed on Convicted Persons by Nature of Indictable Offence
Total and Provinces* – Concluded

TABLEAU 6 Peine additionnelle imposée aux personnes condamnées selon la nature de l'acte criminel
Total et provinces* – Fin

Indictable offence / Acte criminel	Extra sentence – Peine additionnelle							
	Total extra sentence / Total peine additionnelle	Leave / Sursis	Fine / Amende	Restitution and fine / Restitution et amende	Good and bound and to keep the peace and fine / Prison et engagement de garder la paix et amende	Prohibition and fine / Interdiction et amende	Other prohibitions / Autres interdictions	Other penalties / Autres peines
	Cases – Causes							
Assault causing bodily harm – Voies de fait qui causent des lésions corporelles	2	–	–	–	–	1	–	–
Assault on peace officer and obstructing – Voies de fait contre un agent de la paix et entrave	3	–	–	–	–	3	–	–
Common assault – Voies de fait simples	1	–	–	–	–	1	–	–
Cruelty to animals, bestiality, bodily harm – Actes cruels envers les animaux, bestialité, lésions	2	–	–	–	–	–	–	–
Criminal negligence, bodily harm – Négligence criminelle, lésions	1	–	–	1	–	–	–	–
Criminal negligence, break and enter (housebreaking) – Négligence criminelle, vol (effraction et entrée)	1	–	–	–	–	–	1	–
Against the person – Contre la personne	7	–	–	3	–	3	1	–
Breaking and entering a place – Introduction par effraction dans un endroit	6	–	–	1	–	3	–	–
Against property with violence – Contre la propriété avec violence	6	–	–	1	–	3	–	–
Having in possession – Avoir en sa possession	2	–	–	–	–	1	–	–
Theft – Vol	19	–	–	2	–	15	–	–
Against property without violence – Contre la propriété sans violence	21	–	–	4	–	17	–	–
Other interference with property – Autres ingérences concernant des biens	2	–	–	–	–	2	–	–
Other than offences against property – Autres infractions concernant des biens	2	–	–	–	–	2	–	–
Conspiracy – Complot	1	–	–	–	–	–	–	–
Other attempts – Autres offenses	2	–	–	–	–	2	–	–
Public mischief – Méfait public	2	–	–	–	–	2	–	–
Other – Autres	3	–	–	–	–	3	–	–
Criminal Code – Code criminel	30	–	–	6	–	24	1	–
Food and Drugs Act – Loi des aliments et drogues	2	–	–	1	–	1	–	–
Federal statutes – Lois fédérales	3	–	–	1	–	1	–	–
Grand – Total – Général	41	–	–	8	–	31	1	–
Assault causing bodily harm – Voies de fait qui causent des lésions corporelles	3	–	–	2	–	1	–	–
Assault on peace officer and obstructing – Voies de fait contre un agent de la paix et entrave	1	–	–	–	–	1	–	–
Criminal negligence, break and enter (housebreaking) – Négligence criminelle, vol (effraction et entrée)	1	–	–	–	–	1	–	–
Dangerous driving – Conduite dangereuse	6	–	–	–	–	–	3	3
Indictable assault, bodily harm – Assaults à la portée d'une personne du sexe féminin	1	–	–	–	–	1	–	–
Against the person – Contre la personne	13	–	–	2	–	6	4	1
Breaking and entering a place – Introduction par effraction dans un endroit	6	–	–	1	1	3	–	–
Robbery – Vol qualifié	6	–	–	–	3	1	1	–
Against property with violence – Contre la propriété avec violence	12	–	–	1	4	4	1	–

* Excludes Quebec and Alberta – Le Québec et l'Alberta non compris

Source: Catalogue 85-201 Statistics of Criminal and Other Offences 1970



STATISTICS CANADA

Selected Property Crimes, 1974, 1982 and 1984

	Offences per 100 000 population			% change in offences per 100 000 population		
	1974	1982	1984	1974-1982	1982-1984	Overall 1974-1984
Breaking and entering	1 040	1 501	1 421	+44	-5	+37
– residential	524	841	846	+59	+1	+61
– business	373	470	405	+26	-14	+9
Theft over $200	355	1 199	1 212	+238	+1	+241
– from a motor vehicle	107	562	554	+425	-1	+418
– bicycle	16	106	127	+563	+20	+694
Theft under $200	2 045	2 436	2 084	+19	-10	+2
Theft of Motor Vehicle	471	353	305	-5	-14	-18
Fraud	338	481	489	+42	+2	+45
– credit card	55	59	65	+69	+10	+86
– cheques	229	292	297	+28	+2	+30
– other frauds	74	129	127	+74	-2	+72
Total Property Crimes	4,218	5 955	5 607	+41	-6	+33

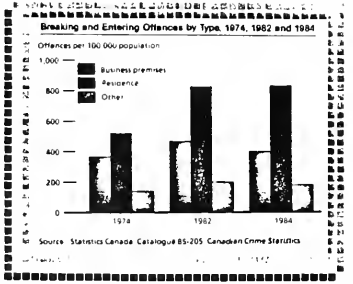
Source: Statistics Canada Catalogue 85-205 *Canadian Crime Statistics*

residential property. Thefts of motor vehicles, on the other hand, declined by 18% in the 1974-1984 period, with most of the decline occurring in the last two years.

Provincial Crime Rates

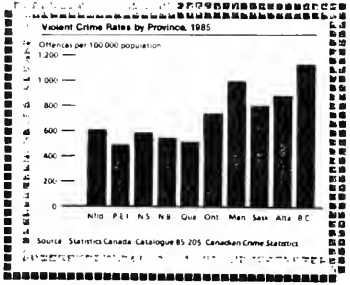
Crime rates in Canada generally increase from east to west across the country. In 1985, for example, British Columbia had by far the greatest incidence of violent crime – 1,135 offences per 100 000 residents of that province. Manitoba had the next highest rate at 998, followed by Alberta with 882. Prince Edward Island had the lowest violent crime rate with just 490 offences per 100 000 persons while Quebec had the second lowest rate at 513.

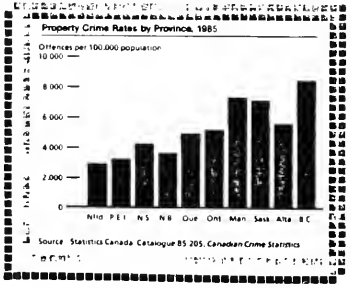
The east to west crime gradient also shows clearly for property crimes. As with violent crimes, the four western-most provinces had the highest property crime rates in 1985 with British Columbia again leading with 8 427 offences per 100,000 population. The four Atlantic Provinces on the other hand had the four lowest property crime rates.



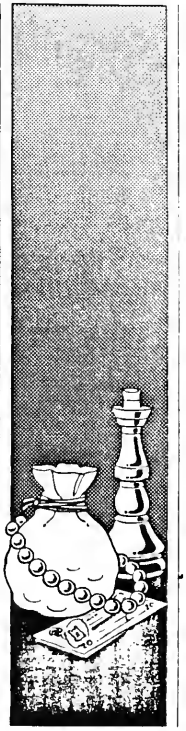
CANADIAN SOCIAL TRENDS - AUTUMN 1986

East-west differences in the incidence of crime are further highlighted in the list of highest and lowest provincial rates for the major types of violent and property crime. Western provinces fill the great majority of the three highest positions in all categories while the three lowest positions are filled almost exclusively by the Atlantic Provinces. Quebec provides the only exception to this trend. On the one hand, Quebec had the highest incidence of robbery of all the provinces and also had relatively high rates of murder, attempted murder and theft of a motor vehicle. On the other hand, Quebec was characterized by low rates of sexual assault and assault. Ontario is the only province not found on either list.





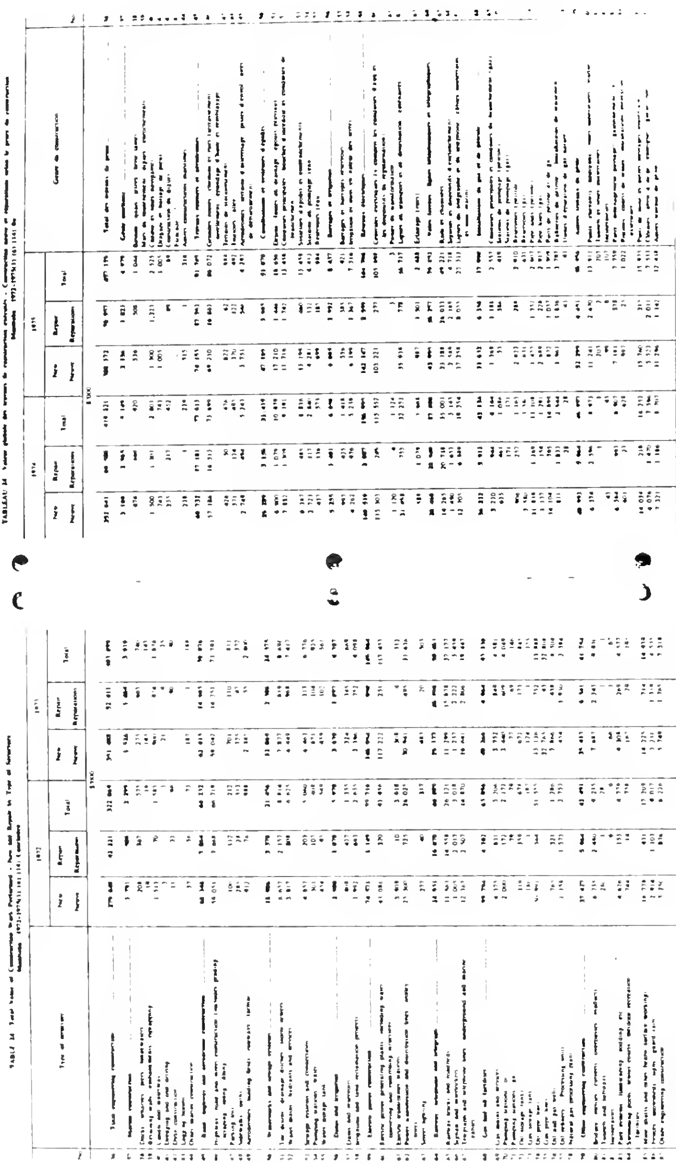
STATISTICS CANADA



CANADIAN SOCIAL TRENDS - AUTUMN 1986

Source: Canadian Social Trends, Autumn 1986

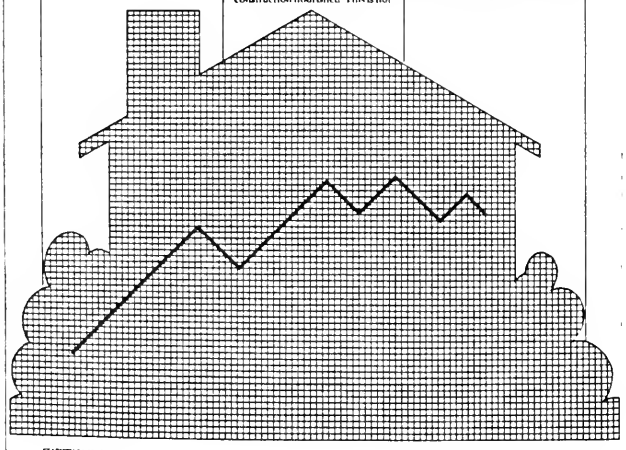
BEFORE



MORTGAGE RATES AND THE HOUSING MARKET

by Alex Berljawsky

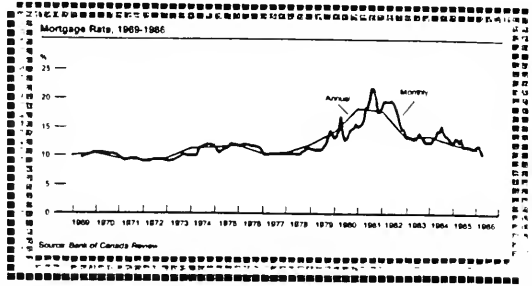
Many Canadians shopping for a house in 1986 were pleased with the real estate market. Mortgage rates were lower than they had been for eight years. As a result, housing sales were brisk, and new construction flourished. This is not surprising given the long term relationship between mortgage rates, house sales and construction activity. People buy, sell and build houses at a livelier pace whenever mortgage rates decline. And the opposite generally occurs whenever mortgage rates go up. The average selling price of existing houses, on the other hand, has generally not been affected by fluctuations in mortgage rates.



Mortgage rates, however, are not the only factor influencing the housing market. Demographic variables such as the rate at which new families are being formed and the size of the potential home-buyer market have considerable effect on housing demand, as do general economic conditions such as the incidence of unemployment and income levels. It should also be noted that national housing market figures mask considerable differences in regional and local housing markets.

Mortgage Rates

Mortgage rates[1] in Canada over the last decade and a half have been characterized by two quite distinct phases. Between June, 1969 and July, 1979 mortgage rates were relatively stable, ranging from 9% to 12%. Throughout the 1980s, however, mortgage rates have been characterized by large and fairly rapid swings. Beginning in the latter part of 1979, mortgage rates escalated rapidly to a peak of 21% in early 1981. In the spring of 1982, rates began a not always smooth descent which saw them fall back to levels similar to those of the 1970's in May, 1986, for example, the five year rate had fallen below 11%.

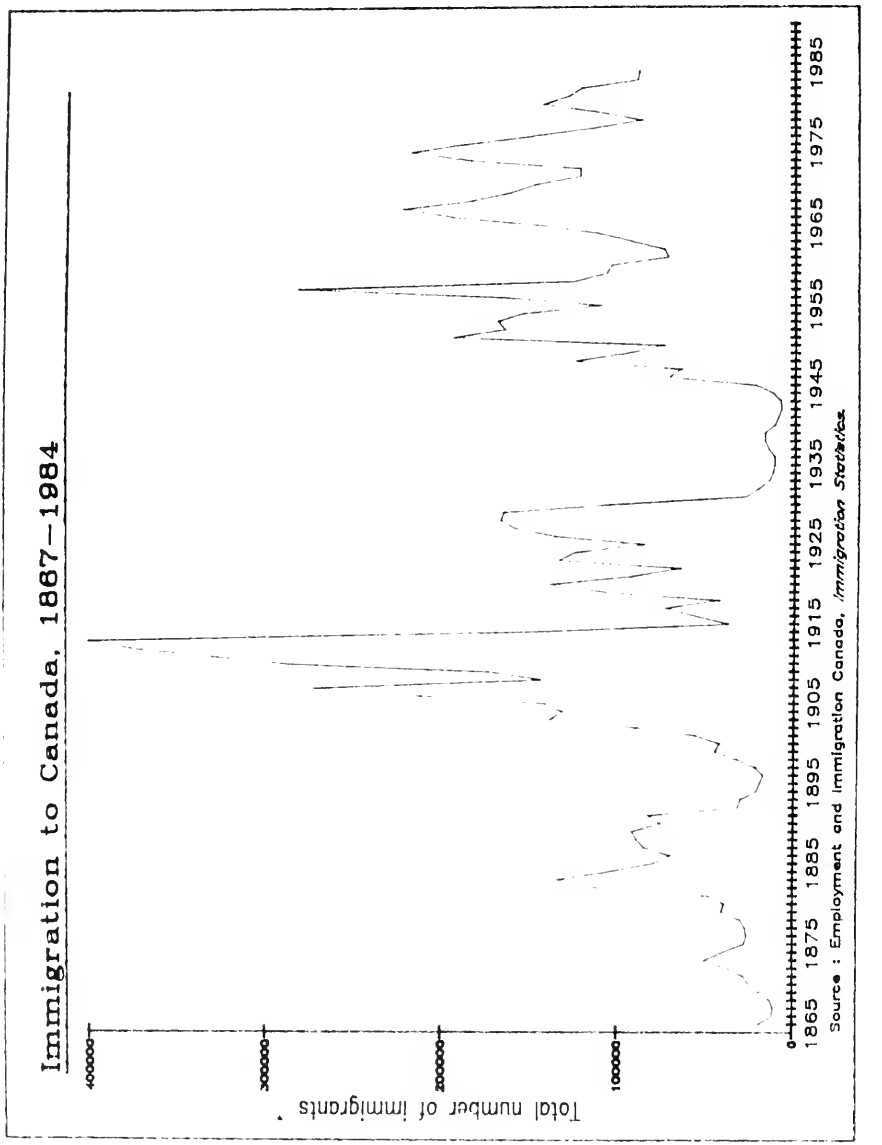


Sales of Existing Houses

The comparison of housing sales[1] with mortgage rates reveals a "braiding" effect. Generally speaking, as mortgage rates rise, housing sales taper off, and when mortgage rates decline, housing sales increase.

This effect was particularly noticeable during the 1980s. During the recession period of 1981-82 mortgage rates escalated rapidly while unit sales dropped in roughly the same proportion. Later as mortgage rates dropped off, unit sales began to rise. The decline in mortgage rates after 1982, for example, was followed by a 50% increase in the number of housing unit sales between 1983 and 1985.

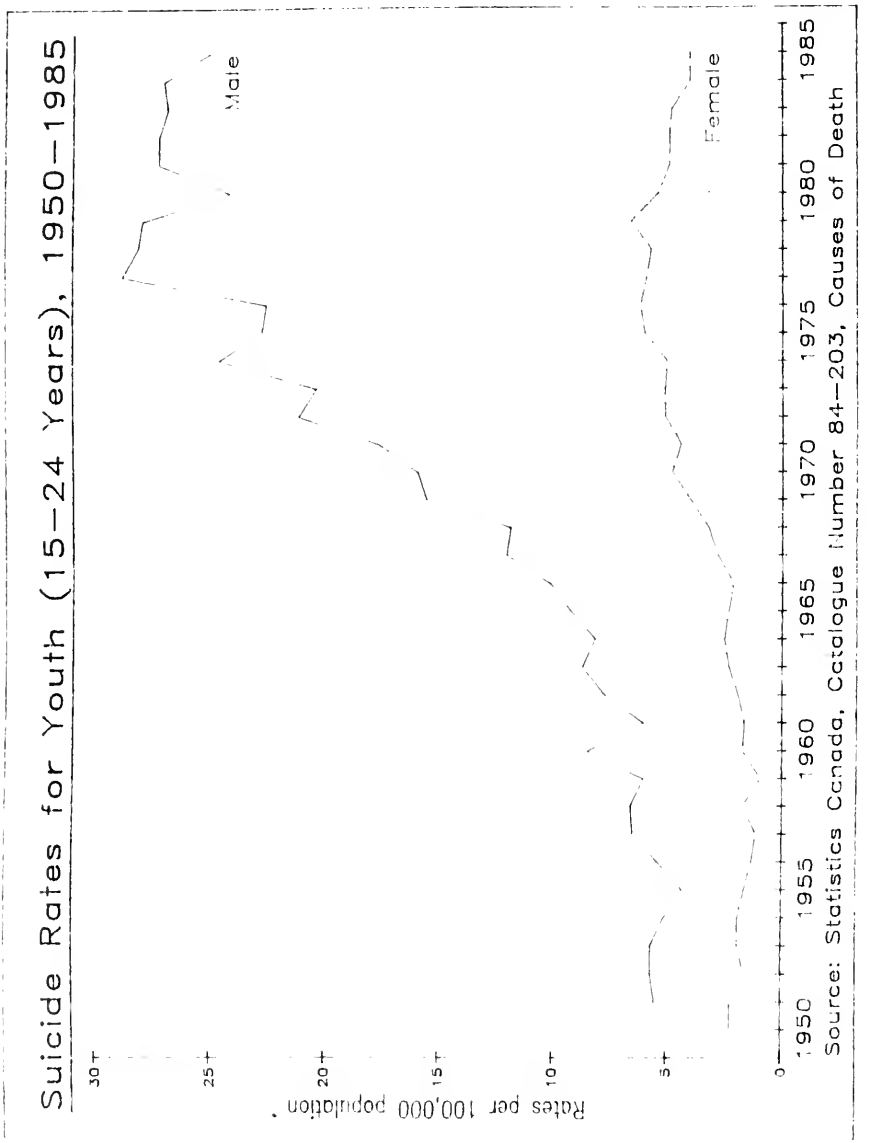
The effect of mortgage rates during these periods of very steep


Immigration to Canada, 1867–1984
Total number of immigrants
400000
300000
200000
100000
0
1865 1875 1885 1895 1905 1915 1925 1935 1945 1955 1965 1975 1985
Source : Employment and Immigration Canada, Immigration Statistics

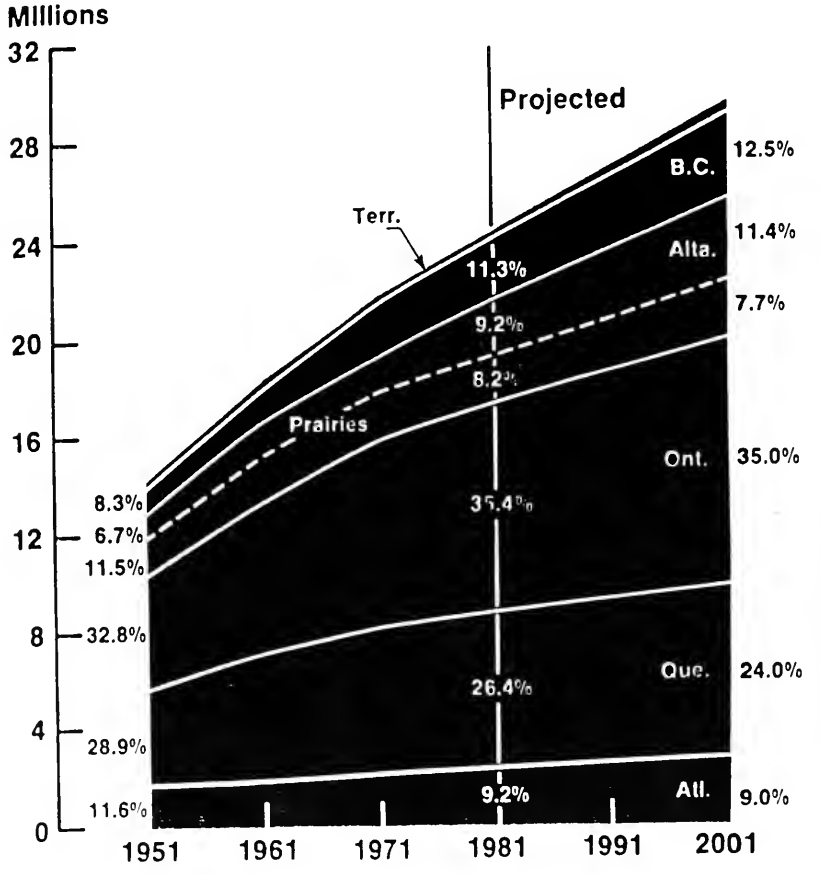
Unemployment Rates by Province

	1975	1983	1985	1986
		%		
Newfoundland	14.0	18.8	21.3	20.0
Prince Edward Island	8.0	12.2	13.2	13.4
Nova Scotia	7.7	13.2	13.8	13.4
New Brunswick	9.8	14.8	15.2	14.4
Quebec	8.1	13.9	11.8	11.0
Ontario	6.3	10.4	8.0	7.0
Manitoba	4.5	9.4	8.1	7.7
Saskatchewan	2.9	7.4	8.1	7.7
Alberta	4.1	10.8	10.1	9.8
British Columbia	8.5	13.8	14.2	12.6
CANADA	6.9	11.9	10.5	9.6

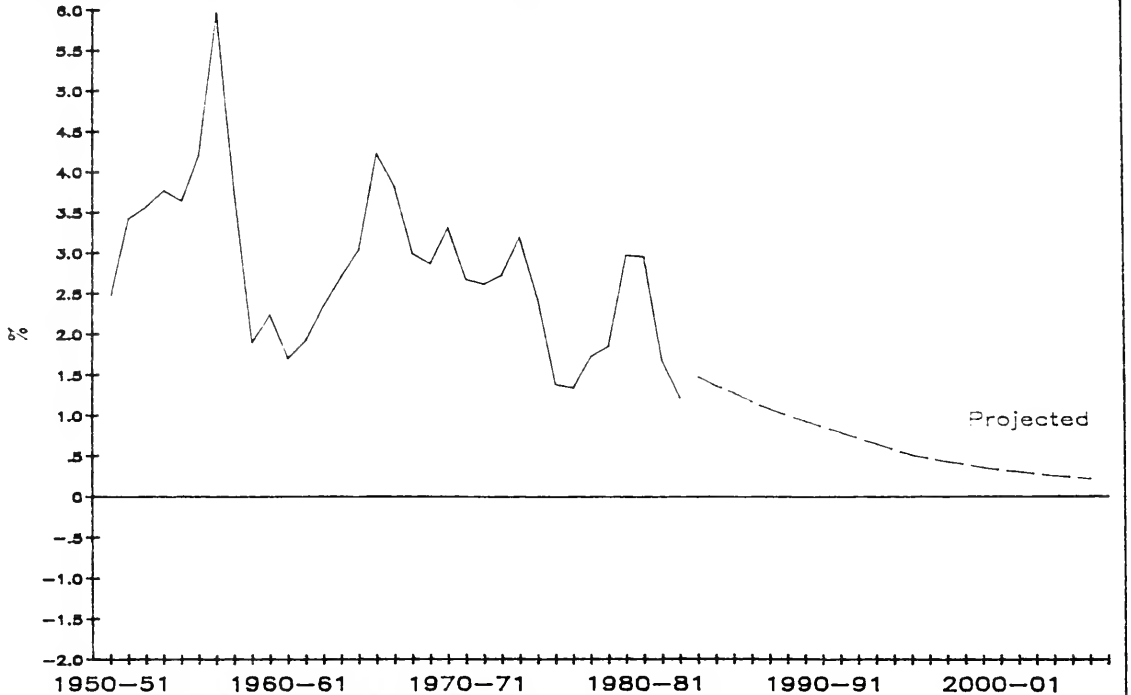
Source : Statistics Canada, Catalogues 71–529, *Labour Force Annual Averages*, and 71–001, *The Labour Force*, December 1986.


Suicide Rates for Youth (15–24 Years), 1950–1985
Rates per 100,000 population
30
25
20
15
10
5
0
1950
1955
1960
1965
1970
1975
1980
1985
Male
Female
Source: Statistics Canada, Catalogue Number 84–203, Causes of Death

Population Distribution by Regions

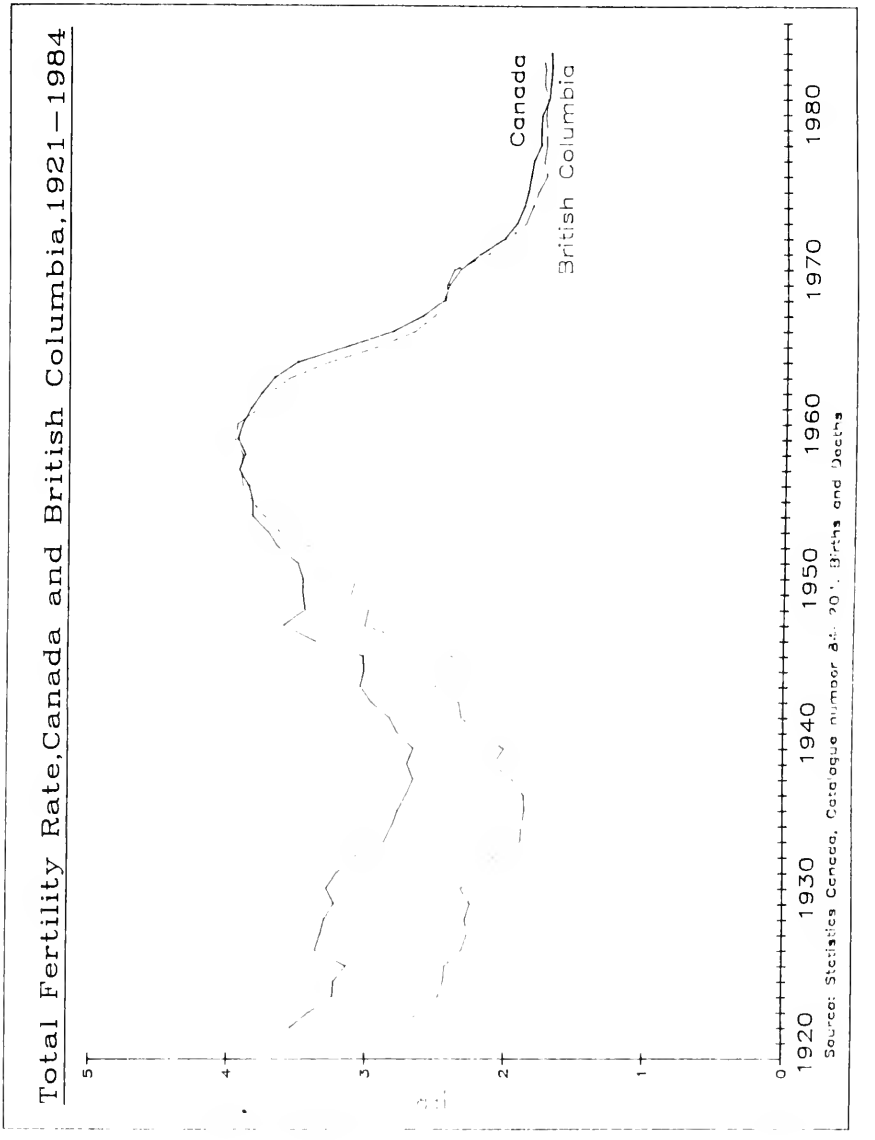


Annual Population Increase for British Columbia, 1950–2006



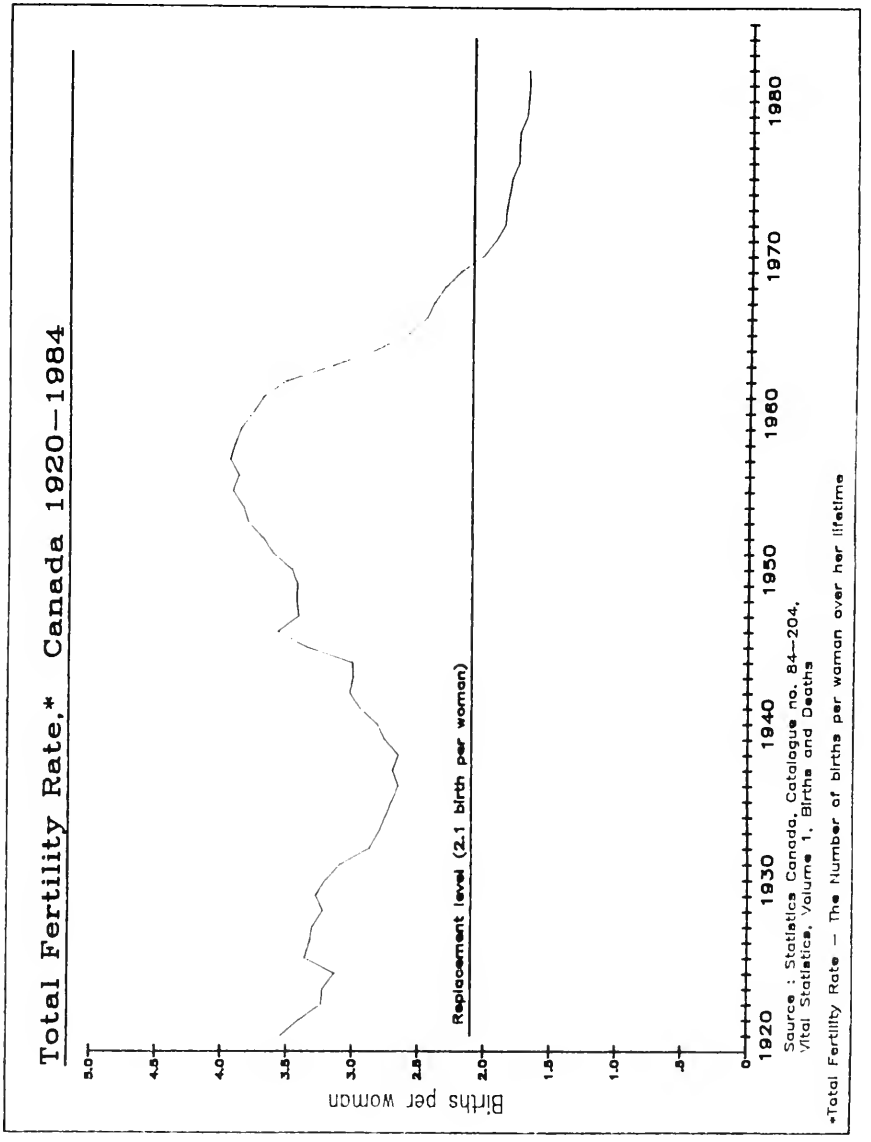
* Projection assumes a total fertility rate of 1.4 births per woman and net migration of 50 000 people.

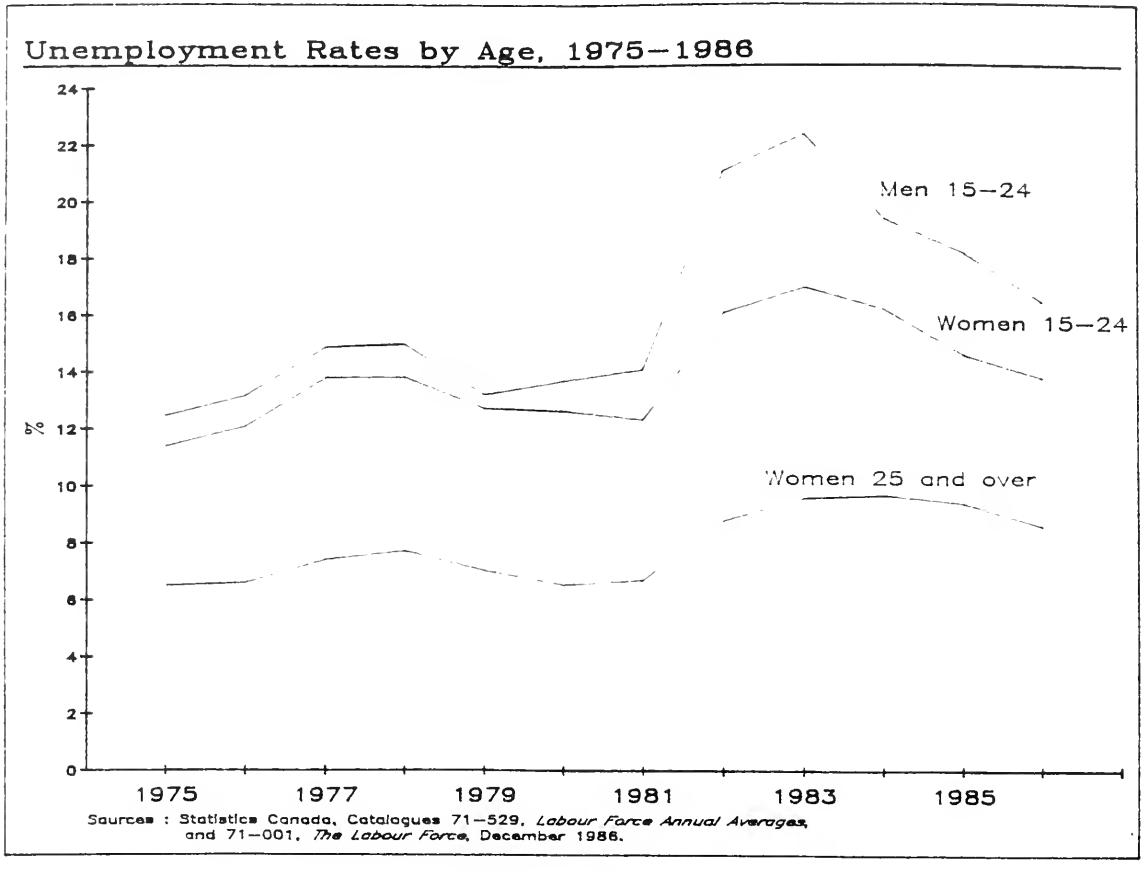
Source: Statistics Canada, Catalogue Nos. 91–210 and 91–520.

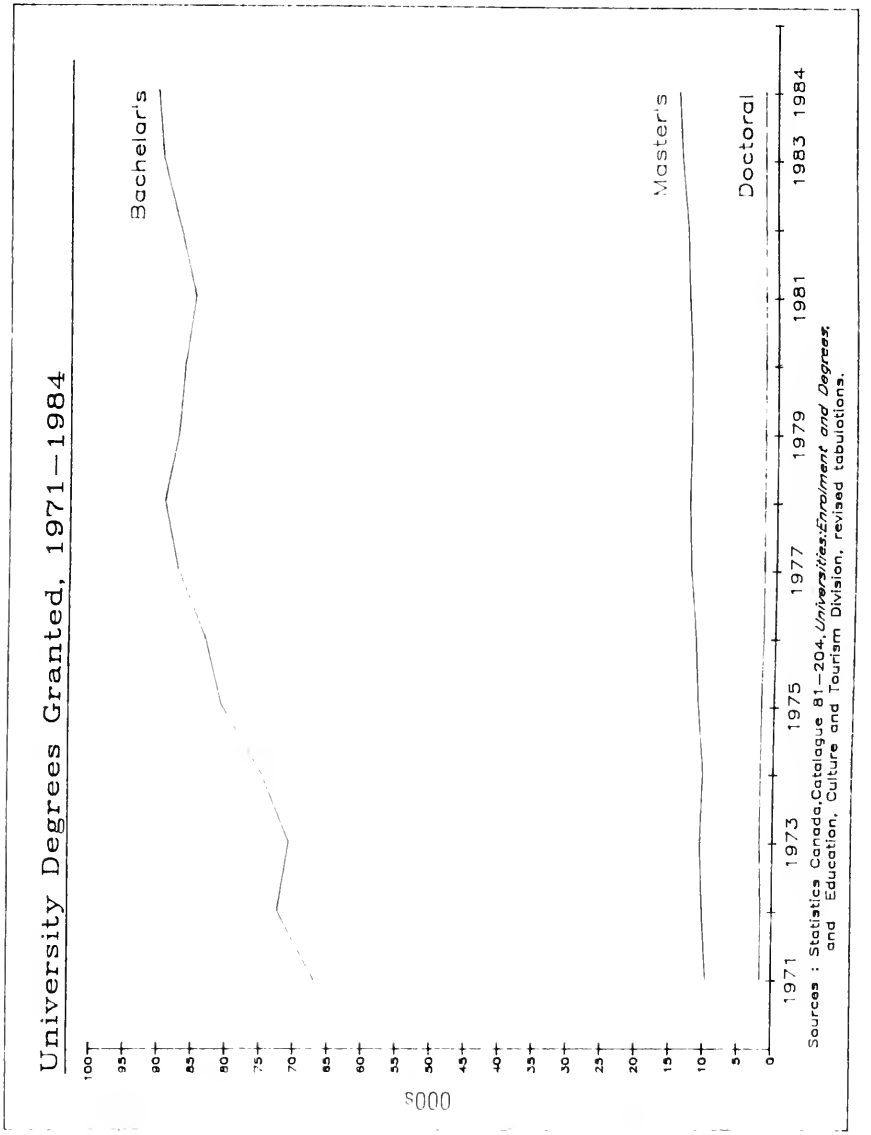

Total Fertility Rate,Canada and British Columbia,1921–1984
Canada
British Columbia
5
4
3
2
1
0
1920
1930
1940
1950
1960
1970
1980
Source: Statistics Canada, Catalogue number 84-204, Births and Deaths

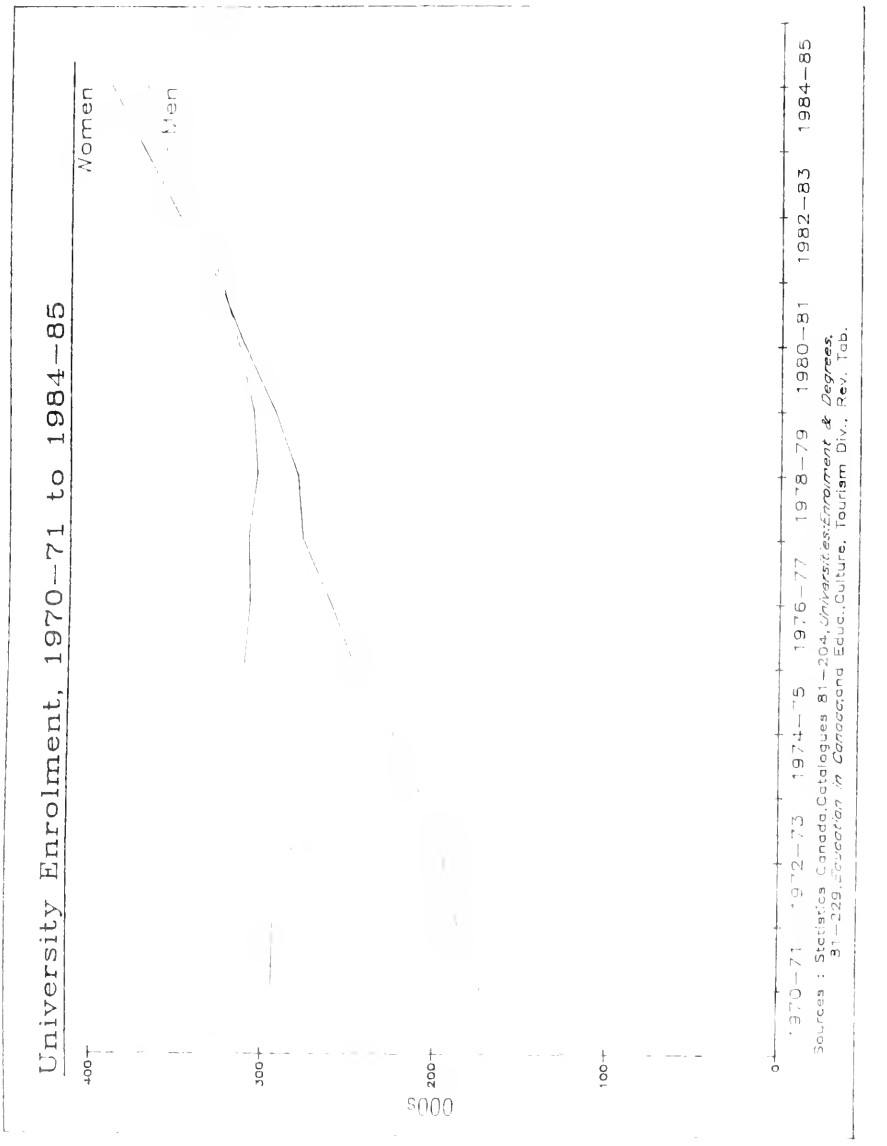
Total Fertility Rate, Canada, Provinces and Territories, 1982–1984
Indice synthetique de fecondite, Canada, provinces et territoires, 1982–1984

	1982	1983	1984
Canada	1.69	1.68	1.69
Newfoundland – Terre–Neuve	1.97*	––	––
Prince Edward Island – Ile du Prince Edouard	1.93	1.89	1.89
Nova Scotia – Nouvelle Ecosse	1.67	1.66	1.63
New Brunswick – Nouveau–Brunswick	1.70	1.69	1.65
Quebec	1.52	1.47	1.46
Ontario	1.65	1.66	1.69
Manitoba	1.84	1.87	1.86
Saskatchewan	2.17	2.13	2.11
Alberta	1.96	1.96	1.92
British Columbia – Colombie–Britannique	1.74	1.73	1.76
Yukon	2.04	2.36	2.25
Northwest Territories – Territoires du Nord–Ouest	3.00	3.20	2.99


Total Fertility Rate,* Canada 1920–1984
Births per woman
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
.5
0
Replacement level (2.1 birth per woman)
1920
1930
1940
1950
1960
1970
1980
Source : Statistics Canada, Catalogue no. 84–204,
Vital Statistics, Volume 1, Births and Deaths
*Total Fertility Rate — The Number of births per woman over her lifetime


Unemployment Rates by Age, 1975–1986
%
24
22
20
18
16
14
12
10
8
6
4
2
0
1975
1977
1979
1981
1983
1985
Men 15–24
Women 15–24
Women 25 and over
Sources : Statistics Canada, Catalogues 71–529, Labour Force Annual Averages, and 71–001, The Labour Force, December 1986.


University Degrees Granted, 1971–1984
Bachelor's
Master's
Doctoral
'000s
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
1971
1973
1975
1977
1979
1981
1983
1984
Sources : Statistics Canada,Catalogue 81-204,*Universities:Enrolment and Degrees*,
and Education, Culture and Tourism Division, revised tabulations.


University Enrolment, 1970–71 to 1984–85
Women
Men
000s
400
300
200
100
0
1970–71
1972–73
1974–75
1976–77
1978–79
1980–81
1982–83
1984–85
Sources : Statistics Canada,Catalogues 81–204,Universities:Enrolment & Degrees;
81–229,Education in Canada;and Educ.,Culture, Tourism Div., Rev. Tab.

CANADA

SOCIAL TRENDS

CONTENTS

Features

Trend Reports

Canadian Social Trends

Editor	David Brusegard
Associate Editor	Craig McKie
Managing Editor	Colin Lindsay
Assistant Editors	Mary Anne Burke, Jo Anne Parliament
Consulting Editor	Lisa Kwong
Art Direction	Greg Moore
Design	Roberto Guido, Bruce Jamieson, Jill Reid
Marketing and Promotion	Judith Buehler, Kathryn Bonner
Composition	Monique Légaré, Rachel Mondou

Acknowledgements

Cynthia Steers, Gordon Priest, Boriss Mazikins, Owen Adams, Doug Angus, Jim MacDonald, Michel Durand, Wendy Hansen, Howard Clifford, Anatole Romaniuk, Sandra Ramsbottom, John Silins, David Bray, Ian Macredie, André Libelle, Sylvie Mercier, Beryl Gorman, Lucie Lamadeleine, Ricarda Windthorst, Georgette Gaulin, Daniel Scott and Sylvie Blais

Cover *Two Girls Reading* Wm. Brymner, watercolour on linen, 40½ X 29¾ in., 1898, National Gallery of Canada, Ottawa

ISSN 0831 5698

CANADIAN

SOCIAL TRENDS

CONTENTS

CANADIAN SOCIAL TRENDS

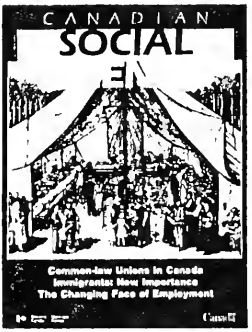
Editor	Craig McKie
Managing Editor	Colin Lindsay
Assistant Editors	Mary Anne Burke, Jo-Anne Parliament
Art Direction and Composition	Publications Division, Statistics Canada
Design	Griffe Design
Photos	Photo Centre, SSC
Promotion	Cherylynn Ireland; Tony Donatucci
Review Committee	J.W. Coombs, J. Hagey, D.B. Petrie, G.E. Priest, E.T. Pryor, M. Rochon
Acknowledgements	Martin Blais, Catherine Bronson, Beryl Gorman, Lucie Lamadeleine, Myriam Laporte, Isabelle Lavoie, Louise Paveley, Cheryl Sarazin, Daniel Scott, Cathy Shea, Tim Stringer

Canadian Social Trends (Catalogue 11-008E) is published four times a year by Statistics Canada. Publication Sales, Ottawa, Ontario, Canada, K1A 0T6, telephone (613) 993-5078.

SUBSCRIPTION RATES: $44 a year in Canada, $50 elsewhere. Single issue $12.50 each in Canada, $15 elsewhere. Send subscription orders and address changes to Statistics Canada, Publication Sales, Ottawa, Ontario, Canada, K1A 0T6. Please supply both old and new addresses and allow six weeks for change. Correspondence may be addressed to the Editor, Canadian Social Trends, 11th Floor, Jean Talon Building, Ottawa, Ontario, K1A 0T6. **Canadian Social Trends** is not responsible for unsolicited materials.

Cover: *Christie Passage, Hurst Island, B.C.* by Edward J. Hughes
Oil on canvas, 1982. © Dominion Gallery, Montreal
Collection: National Gallery of Canada, Ottawa

ISSN 0841-5698



SOCIAL TRENDS

CONTENTS

CANADIAN SOCIAL TRENDS

Editor	Craig McKie
Managing Editor	Colin Lindsay
Assistant Editors	Mary Anne Burke, Jo-Anne Parliament
Art Direction and Composition	Publications Division, Statistics Canada
Design	Griffe Design
Photos	Regional and Industrial Expansion, Photo Centre, SSC
Review Committee	J W Coombs, J Hagey, D B Petrie, G E Priest, E T Pryor, M Rochon
Acknowledgements	Sylvie Blais, Lucie Lamadeleine, Elizabeth Marcella, Kate McGregor, Suzanne Méthot, Louise Quinn, Sandra Ramsbottom, Daniel Scott, Cathy Shea, Tim Stringer

Canadian Social Trends (Catalogue 11-008E) is published four times a year by Statistics Canada, Publication Sales, Ottawa, Ontario, Canada, K1A 0T6, telephone (613) 993-5078

SUBSCRIPTION RATES $44 a year in Canada, $50 elsewhere. Single issue $12.50 each in Canada, $15 elsewhere. Send subscription orders and address changes to Statistics Canada, Publication Sales, Ottawa, Ontario, Canada, K1A 0T6. Please supply both old and new addresses and allow six weeks for change. Correspondence may be addressed to the Editor, **Canadian Social Trends**, 11th Floor, Jean Talon Building, Ottawa, Ontario, K1A 0T6. **Canadian Social Trends** is not responsible for unsolicited materials.

Cover: ***Manitoba Party*** by William Kurelek
oil and pencil on board 48" x 60" 1964
© (Mrs. William Kurelek) The Isaacs Gallery, Toronto
Collection National Gallery of Canada, Ottawa
ISSN 0831-5698

The *New OED* project at Waterloo:
old wine in new bottles

by D. W. Russell[1]
Waterloo Centre for the New OED
University of Waterloo

In mid-1983, three years before the fourth and final volume of the *Supplement to the Oxford English Dictionary* (OED) appeared in conventional print form, Oxford University Press (OUP) had announced its intention of computerizing the *OED*. By the end of 1983 the decision had been made that Oxford University Press would manage this large undertaking itself, and would enter into a series of contracts/agreements with other parties (such as the University of Waterloo, IBM UK, International Computaprint Corporation (ICC) of Fort Washington, Pennsylvania, among others) to carry out work on various aspects of the computerization.

From a long-term perspective, the Project's objective is to transform both the original 12 volumes of the *OED* and the 4-volume *Supplement* into an electronic database; the work to be done in arriving at this objective has been broken down into several discrete phases: first, the initial sixteen volumes had to be entered into the computer, preserving the original text organization and presentation. This meant the tagging of structural and typographical elements in the dictionary as the data were keyed and transfered onto magnetic tape. One of the aims of this phase is to produce a printed version of the dictionary, integrating the *Supplement* material with the body of the *OED*, and incorporating about 4000 new entries into this merged edition. This version will be printed in the spring of 1989, in a projected 22 volumes at a price of about £1,500. The next major phase involves the design of a database structure for the machine-readable data, so that alternative structures can be presented, and interactive querying will be possible. It is at this point that the *New OED* comes into being, leading to expanded, updated, and revised versions of the dictionary that will allow several modes of data access, ranging from direct, online access, to the conventional printed version, and special, printed subsets of the dictionary. As can be seen, the possibilities are many, and the prospects for

[1]Presented at the International Association for Social Science Information Service and Technology (IASSIST) Conference held in Vancouver, British Columbia, Canada on May 19-22, 1987

lexicographical work seem to expand almost infinitely.

The magnitude of the proposed task can be appreciated when one considers the size of the current source data, the sixteen volumes of *OED* and *Supplement*: there are over 21,000 pages of three-column print, with a total of about 500 million characters including punctuation and spacing. This breaks down to about 306,000 main entries, 163,000 subordinate entries, with over 2,350,000 illustrative quotations, and almost half a million cross references. In addition to the sheer size of the dictionary, the material has, as anyone who has ever used the *OED* knows, an extremely complex structure, by which explicit and implicit information is conveyed to the reader through both the layout and the typography of the text. At the simplest level, each entry usually includes a headword lemma, a pronunciation key, a grammatical category label, a list of variant forms given by century, an etymological section, a series of sense definitions, each with a quotation bank; there is usually one quotation per century, with date, author, source and bibliographic reference. This structure is made more complex by words or forms that do not fit easily into the usual categories, and by the inconsistencies in method inevitable in a project that spanned many years and several generations of lexicographers.

Still, the capture of the dictionary material, which had to be done by manual keying of the text, since it was too complex for optical scanners, has been done, and done surprisingly successfully over a period of 18 months, ending in June 1986. The error rate, as noted by human proof readers hired by Oxford, is seven or fewer errors per 10,000 keystrokes. Working from enlarged copies of *OED* text, the key entry operators entered tags to identify all the typographical elements, as well as some of the structural elements of the source text. One of the chief aims at this stage, was, after all, to be able to produce a new, typeset edition in 1989.

But, since even this seemingly routine task becomes more complex when one is dealing with as much material as is found in the *OED* and *Supplement*, a further refinement was introduced at this point in the process. Computer science researchers at Waterloo developed a parser which automatically tagged structural elements not tagged during keying, and which converted the ICC tags into SGML codes.[2] This parsing permitted the automatic validation of the earlier tagging. The conversion to SGML codes was needed to permit a greater degree of automatic integration by computer of the *OED* and *Supplement*, as well as to facilitate the lexicographical team's interactive integration of the two source texts.

Meanwhile, at the University of Waterloo, research on database design for the *New OED* has begun, with the support of a 1.3 million dollar grant from NSERC. The Project has the mandate to design a database for the *New OED*, and to develop software utilities for database access, maintenance and update. This work is to be carried out over three years by a team of seven full-time researchers, directed by a team of computer science faculty researchers, led by Gaston Gonnet and Frank Tompa. To date, two prototype software tools have been produced, to allow interactive access to, and sophisticated querying of, the database. These tools are named *Goedel* and *Pat*. Although it is not within my competence to describe any of the technical specifics of these tools, I would like to offer some examples of the results currently possible, drawing mainly from my own research interests in the *OED*, namely the identification and study of the Anglo-Norman elements which were adopted by the English language in the medieval period.

[2]See Rick Kazman, *Structuring the text of the Oxford English Dictionary through finite state transduction*, (M. Math. thesis) University of Waterloo, 1985.

When the first tapes of the *OED* became available at Waterloo in 1985, it was theoretically possible to begin searching the dictionary interactively. In order to find all words labelled as Anglo–French or Law French, I needed simply to ask the computer to list occurrences of the relevant strings, "AF", "ONF", "Law Fr.", etc. At the time the data were mounted in a series of files, which meant the queries had to be repeated for each file. The results were not easy to read, and although one could scroll back or forward to identify the headword, in the end it proved easier to use the printed dictionary to locate the relevant entry.

After the creation of *Goedel* in 1986, a whole new range of possibilities made my querying of the data easier. I could now ask for the extraction of material according to structural categories in the dictionary entries, and according to specific strings within each category. I decided to limit my extraction to listing the headword lemma, the material within the etymological section, and the dates of the illustrative quotations, for all entries which could be considered to be derived from Anglo–Norman sources. Within the structure of the *OED*, Anglo–Norman material is identified in various fashions in the etymology section: it may be labelled as Anglo–French, it may be labelled as Old French (or Old Norman French or Old Law French), or it may be found to be labelled as French but with quotation dates preceding 1500. The results of my queries using *Goedel* could be printed in a formatted form which is eminently readable, and could be carried away for use elsewhere, freeing the researcher from being tied to a computer terminal. With the extraction power and flexibility of *Goedel*, the humanist researcher is faced with the new challenge of creating more sophisticated queries, based on previously unexamined possibilities, and building on the results of his or her ongoing interactive research.

The second extraction tool, *Pat*, allows a rapid and effective querying of a different sort, based in part on pattern matching; with *Pat*, for example, I could extract all entries derived from AF or OF and which have supporting quotations from a particular author, such as Chaucer or Gower. Or, to give another example, a researcher looking for infantine language was able to search the dictionary for all words whose sense definition included specific key words, such as "little" followed by "boy" or "girl" or "child" within a specified number of characters. Increasing familiarity with the results of these searches led to more sophisticated querying of the database, and raised technical problems which were addressed by the data structuring group working on the Project. In a similar way, *Pat* was used by a researcher interested in the source and frequency of quotations. It is possible to extract all quotations from a particular author, such as Fanny Burney, and further, to extract only quotations from Burney from a specified work, such as *Cecilia*.

These short examples, from among the preliminary group of research projects underway at the Waterloo Centre, serve to emphasize the futuristic nature of the Project; it is difficult to design a database for uses which may arise in the future, but which have not yet been imagined by humanist researchers. In an attempt to come to grips with this problem, OUP and the University of Waterloo conducted a user survey, to find out how individuals use the *OED*, to determine the principal facilities needed for the *New OED*, and to provoke considered responses about applications for the electronic version of the dictionary. Over 1,100 individuals were contacted, of whom 60% were from the UK and Europe, 40% from North America. The sample included both academic and non–academic users, with an emphasis on sophisticated users. The results have not yet been completely analyzed, but preliminary results give a broad outline of what users would like to see in the *New OED*. Basically, most

users want everything currently in the *OED*, in a fashion that is simple to use, quick to give results, and cheap to access. Ideally the data will be available publicly through an online data service, and privately via some disk format. The software must be user friendly, allowing quick access, both to a skeleton summary of each entry, and to a complete entry or selected details of an entry. The electronic version must not be prohibitively expensive, and yet be amenable to continuous updating and revision. And finally, the *New OED* should continue to be published in printed form. The survey results also suggest future applications, many of which will require revision of the *OED*: these include semantic field searches, frequency ratings by date or language of origin, the use of the *OED* as a thesaurus, and so on. One of the suggested applications, requiring the ability to search phonetics elements, will be made much easier by the decision to replace Murray's phonetic transcriptions with transcriptions based on the symbols used by the International Phonetic Association. This revision is to be incorporated in the 1989 print version of the *New OED*. Searching on these phonetic elements will only be possible, of course, in the electronic version of the *New OED*.

Just what the electronic form of the *New OED* will be has not yet been decided. There are plans to market an exploratory electronic issue of the old *OED*, without the *Supplement*, and without revisions, on CD-ROM in late 1987, with the aim of validating some of the results of the user survey, and generating feed-back for the database design currently in progress for the *New OED*. The database created to produce the print version of the *New OED* in 1989 is not now in a form which OUP would offer to outside users, but the Project does expect to market an electronic version of the 1989 and later editions of the *New OED*. The undertaking does present a number of technical and legal problems, the solutions to which will determine the final end product. And the planned revision and enhancement of the *New OED* after 1989 will certainly carry the Project well into the twenty-first century. Judging from the unforeseen shifts and changes in plan which plagued James Murray and the other editors involved in the creation of the *OED* from 1879 to 1933, it is perhaps unwise to promise completion of the *New OED* Project by any given date. But there is room for cautious optimism. Such is the position taken by Tim Benbow, OUP's Director of the *New OED* Project, in his Status Report given at our second annual conference at Waterloo in November 1986. He said:[3]

> "At the moment the project is running on schedule and on budget. We are not, however, complacent. Recurrent nightmares — one, in which as a juggler one is constrained to keep an increasing number of dictionary volumes of monstrous size in the air under pain of lexicution, alternating with a Sisyphean vision of ordering acres of dictionary slips only to have them taken by the wind as the last is about to be positioned — see to that!"

It is, no doubt, significant that Tim Benbow's nightmares do not yet reflect the presence of any textual database monsters.□

[3]Tim Benbow, "Status Report on the *New OED* Project." Paper given at "Advances in Lexicology", Second Annual Conference of the UW Centre for the *New Oxford English Dictionary*, Waterloo, Nov. 9-11, 1986.

The sampling bias in random digit dialing

by A. Dianne Schmidley[1]
Bell Atlantic Corporation

All research is based on information from data collection efforts, whether the data result from the use of qualitative approaches, such as, content analysis, focus group sessions, or in-depth interviewing by psychologists, social workers or ethnographers, or from methods which produce data more amenable to quantitative analyses, such as, sample surveys and censuses (technically, a census is a 100% sample survey). A study[2], conducted by Derek Phillips in the early 1970s, concerning the preference of sociologists for either quantitative or qualitative data collection approaches, revealed that more than 90 percent of the research conducted by those social scientists resulted from the analysis of data collected through the administration of interview schedules and/or questionnaires developed for use in a survey setting.

While sociologists are generally viewed as major collectors and users of the data resulting from surveys, they are not the only individuals interested in survey data. Pollsters, advertisers, program administrators and others have important uses for survey data. In fact, there are several broad categories of surveys including:

1. Attitude and public opinion polls,
2. Marketing research, including advertising and public relations surveys,
3. Government surveys, especially those conducted for the purposes of developing legislation and administering and evaluating the effectiveness of programs,
4. Special surveys conducted by researchers in university settings, usually the basis of primary research, and often government funded,
5. Other surveys, including internal

[1]Presented at the International Association for Social Science Information Service and Technology (IASSIST) Conference held in Marina Del Rey, California on May 21-24, 1986

[2]Knowledge from What?

organizational surveys conducted to monitor attitudes and opinions.

Individuals with training in survey research methods direct many of these collection activities; however, many surveys, particularly those in the areas of marketing, public relations and opinion research, are conducted by individuals who have no formal training and are not aware of the many pitfalls of survey data. Once the ill-gotten information is translated into copy, it acquires a life of its own and the 'facts' become almost impossible to erase.

The use of data from a survey, for some purpose other than that for which it was collected, is common. For example, data from government surveys such as the Current Population Survey (CPS) are often used as the basis of marketing decisions, while data from a public opinion survey may influence the decisions of government policy makers.

The data from all five categories of surveys are utilized to support policy decisions affecting the expenditure of millions of dollars. Given the penchant of policy makers for basing decisions on data collected through the use of surveys, the need for a critical appraisal of each of the various stages of the survey process has evolved. Questionnaire/interviewer schedule design, sample selection, the administration of the survey process, the collection, tabulation and interpretation of the data, and testing of the reliability and validity of the information collected from the survey, compared to some benchmark, have all become subareas of the survey research process and are carefully scrutinized by trained social scientists.

The purpose of this paper is to shed additional light on the problem of selecting a representative sample of a population to be surveyed, using the procedure of simple random sampling.

When a sample is selected for a survey, attention must be given to the parameters of the universe the researcher hopes to study, and the method of eliciting information from that universe. In the case where the universe to be studied is a human population, there are a limited number of ways of operationally defining the universe and collecting information from the individuals comprising that universe. Each type of definition will affect the method of sample selection chosen. If the universe is operationally defined by street address records, the researcher can mail a questionnaire to each individual respondent. If telephone numbers constitute the universe, the researcher can telephone respondents and administer a questionnaire/interview schedule. With street address records, the researcher can meet with each respondent and conduct interviews face to face.

Generally, evaluations of each of these three approaches suggest that the higher the quality of the data collected (given proper development of the survey instrument), the greater is the expense of the survey. Mail-back questionnaires are the least expensive way in which to collect information and result in the poorest quality data with the lowest response rates, while in-person interviews are the most expensive way in which to collect the information, but produce better respondent cooperation. Given this set of circumstances most researchers employ the telephone as a means of collecting survey information, since it produces medium quality data at medium cost.

I mentioned earlier that when drawing a sample from a universe, the researcher operationalizes the sampling process by giving the members of the populaton s/he hopes to study concreteness in the form of a street address or a telephone number. Client/customer specific lists, collected by some agencies and firms, can be used when the researcher is interested in some subset of the population, and thereby has access to the specific names, telephone numbers and/or

address records of that group.

In most instances, however, the universe to be studied is only vaguely known. It may be comprised of all the individuals in some specific geographic location such as a trading area, county or school zone, or it may be a subset of the population, such as households with a certain income, or people of certain ages or ethnicity. At any rate, owing to a lack of specific information which could be used to contact and interview the individuals of interest in the population, the researcher more often uses a method of simple random sampling (SRS) of available street records or telephone numbers in order to delineate those units which will be sampled for the purpose of collecting information which can then be ascribed to the larger universe from which the sample was drawn.

The preference of survey researchers for telephone interviewing as a means of collecting information from respondents, coupled with the need to employ SRS in order to to identify the individuals to be surveyed within the universe in which they are located, has led recently to the development of a technique called 'Random Digit Dialing' (RDD).

RDD is believed by the naive to be a cure for drawing biased or non-representative samples from a universe. Misinformed advocates of RDD persist in believing that universal telephone service means that everyone has a telephone in their home. Secure in this belief, RDD practicioners promote the idea that by simply dialing the telephone in a random manner one can draw a random sample of the population of any given geographic area. Census data, which are used by experienced survey researchers, to design sampling frames and to calibrate survey results, indicate that universal telephone service does not mean that a telephone is found in every housing unit.

'Universal telephone service' is akin to the notion of 'full employment'. Full employment is defined by many as the situation in which about 5% of the population in the labor force, ages 16 through 64, and desiring employment, is unemployed. There are many persons between the ages of 16 and 64 who are not employed, and are not seeking employment; so full employment does not mean that 5% of this entire age group is unemployed. Similarly, universal telephone service does not mean that every housing unit contains telephone network access. There are a variety of reasons why a unit may be without access to the telephone network.

Attachment A to this paper gives the reader the sources utilized to determine something called telephone penetration. As can be seen in this attachment, there are basically four ways in which to measure telephone penetration: (1) using decennial census data, (2) using information from the CPS, (3) using GBF-DIME records, and (4) comparing household counts with counts of telephone access lines in given geographic areas.

Examples of the decennial census and CPS measures of telephone penetration of each of jurisdictions served by Bell Atlantic appear in Table 1, entitled, "Telephone Availability For Selected Areas: United States and Bell Atlantic Served States, 1980-1985". Figures 1 and 2 illustrate changes over time in telephone penetration in these geographic locations. Table 2 provides and indication of the type of statistic that can be developed using the Access Lines/Households ratio.

Several conclusions are possible from a careful examination of these various measures of telephone penetration:

1. Data from different questions result in different statistics; i.e., the two censuses produced different results partly owing to question wording. These differences,

although found in a comparison of data from the 1970 and 1980 censuses, are illustrated in Table 1 in the 'unit' and 'availability' measures from the CPS,

2. In addition to these differences, however, there are major differences between geographic locations (Figures 1 and 2) and,

3. Differences have occurred over time (Table 2),

4. Differences result from the sample selection process. Table 1a reveals the differences in sample drawn from 1980 census data (Columns a and b); Table 1b illustrates the range associated with each confidence interval computed for the various point estimates of penetration computed from CPS data.

Thus far, we have examined the penetration rate for the total number of households. When differences in types of households is taken into account, the variation in telephone penetration becomes even greater. The data in in Table 3 are from a 1980 Census Public Use Microdata Sample (PUMS). Table 3 contains a cross-tabulation of households by age of householder, household income, and telephone penetration for the state of Virginia. These data indicate that the presence of a telephone in the housing unit is not a random event, but rather that one is less likely to have a phone if one is poor and/or young.

Figure 3 illustrates the ratio of telephone access lines to households in the Bell Atlantic region from the early 1950s to the 1990s forecast period. A surprising event occurred in the late 1970s: the number of access lines increased to the point of outnumbering households. How can this be? The answer is quite simple, it's called multiple lines. In the state of New Jersey alone, more than 10% of households have at least two lines, and these households are not poor and they are not young.

Can one draw a random sample, utilizing RDD? Yes, but a random sample of what? Certainly one will not draw a random sample of the households in a given geographic location. One will draw a random sample of telephone lines and, as we have seen, these are not evenly distributed across the population.

Does this matter? It depends on what one is trying to accomplish. A pollster, trying to predict an election, will want to examine age specific voting patterns and calibrate these to the universe of households with telephones, allowing for multiple lines, of course. Since the people who vote tend to be those with higher incomes, and voter participation increases with age, you will reach the group you wish to sample using telephone interviewing, although you may overstate the case, especially if you are tracking Republican candidates. (Table 4)

A marketing researcher, must be careful to calibrate the resulting data with census data, because the upscale population will be over-represented and other groups under-represented in the sample. An academic or government researcher, conducting a survey designed to produce information for the development of legislation or the operation of a social program, would need to calibrate the data with census or CPS data in order to ensure a representativeness.

In conclusion, it seems highly unlikely that telephone interviewing is a reasonable replacement for the Bureau of the Census traditional decennial census methods of data collection employing a housing address list, mailout questionnaires and enumerators. Given the problems associated with the universe of telephone access lines, if telephone interviewing becomes the mainstay of data collection for the U.S. Government, we will have lost one of the most important means we have of analysing the composition of our population. ¤

Attachment A

METHODS OF MEASURING TELEPHONE AVAILABILITY/PENETRATION (STATE LEVEL)

I. Decennial Census of Population and Housing

Questionnaire

1970 #H1 Is there a telephone on which people living in your quarters can be called?

☐ yes what is the number?

\- - - - - - - - - - -
Phone number

☐ no

1980 #H26 Do you have a telephone in your living quarters?

☐ yes ☐ no

Caveats:
- Timliness of the data
- Validity & Reliability issues

II. Current Population Survey

Interview

Supplementary questions regarding telephone availability asked in March, July & November of each year of a state based sample of households

Caveats:
- Sampling error at state level
- Coverage of non-Bell Atlantic served areas

III. Geobased Files/Dual Independent Map Encoding

Matching telephone customer address records with street address records in the GBF

Caveats:
- Non-urbanized areas not covered
- New housing not covered, although it can be inferred from telephone company records

IV. Residence Acess Lines/Household Estimates;

Use telephone company access line information in the numerator and household estimates prepared by company demographer in the denominator

Caveats:
- Second line development; failure of business office to identify second line
- Accuracy of household estimates; OK for internal purposes, may not stand up in legal/regulatory setting

Table 1-A

TELEPHONE AVAILABILITY FOR SLECTED AREAS: UNITED STATES AND BELL ATLANTIC SERVED STATES, 1980-1985

GEOGRAPHIC UNIT	1980 (A)	1980 (B)	NOV 1983		MARCH 1984		JULY 1984		NOV 1984		MARCH 1985		JULY 1985		6 MONTH AV. (1-6-85)	
			Unit	Avail.	Unit	Avail.	Unit	Avail.	Unit	Avail.	Unit	Avail.	Unit	Avail.	Unit	Avail.
United States	92.9	92.9	91.4	93.7	91.8	93.6	91.6	93.8	91.4	93.6	91.8	93.7	91.8	93.9	91.6	93.7
Washington D.C. (C&P)	96.0	95.3	94.7	95.6	96.1	97.5	93.5	95.4	95.1	96.0	91.6	93.5	93.6	94.9	94.1	95.5
Maryland (C&P of MD)	96.0	95.8	96.3	96.7	96.1	96.9	94.9	95.7	96.1	96.8	95.2	96.2	96.2	98.1	95.8	96.7
Virginia (C&P of VA)	92.6	91.8	93.1	94.7	93.2	95.1	93.0	95.6	92.9	94.6	92.8	94.5	90.4	92.3	92.6	94.5
West Virginia (C&P of WVA)	89.6	89.3	88.1	91.1	87.2	93.5	86.5	90.0	89.4	92.1	88.1	91.4	88.7	92.8	88.0	91.8
New Jersey (NJ Bell Tel.)	95.4	95.3	94.1	95.1	93.5	95.0	96.0	96.9	94.5	96.3	95.1	96.5	95.4	96.5	94.8	96.1
Pennsylvania (Bell of PA)	96.0	95.9	95.1	97.1	94.4	96.0	95.1	96.4	95.1	97.2	94.2	95.5	95.8	96.8	94.9	96.5
Delaware (Diamond State)	95.2	95.1	95.0	96.6	95.4	96.3	93.7	95.1	93.7	95.8	96.6	97.4	94.4	96.1	94.8	96.2

Sources: 1980(A): PHC 80-S1-1 "Provisional Estimates of Social, Economic and Housing Characteristics", Table #1; 1980. (Official numbers files by AT&T with the DOJ, RE. The Modified Final Judgement, 1982.)

1980(B): STF3 - Final Numbers from the 1980 Census of Population & Housing.

Nov. 1983 - July 1985: Current Population Survey (CPS)

Table 1-B

Confidence Intervals Associated With Point Estimates
Of Telephone Penetration for Bell Atlantic Served Jurisdictions

Prepared by D. Schmidley
Demographic Studies
x8638

(1) State	(2) Unit Measure Nov. 1985	(3) CVx	(4) SEx	(5) Range 68%	(6) Range 95+ %
DC	95.6	0.0149	1.42	94.2 - 97.0	92.8 - 98.4
DEL	93.4	0.0119	1.11	92.3 - 94.5	91.2 - 95.6
MD	95.3	0.0120	1.14	94.2 - 96.4	93.0 - 97.6
NJ	94.1	0.0093	0.88	93.2 - 95.0	92.3 - 95.9
PA	95.8	0.0059	0.57	95.2 - 96.4	94.7 - 96.9
VA	92.0	0.0156	1.44	90.6 - 93.4	89.1 - 94.9
WVA	86.1	0.0183	1.58	84.5 - 87.7	82.9 - 89.3

Formula from U.S. Bureau of Census

$$SE_x = X \cdot CV_x$$

X1 = 68% Confidence Interval

X2 = 95% Confidence Interval

SE_x = Standard error

X = Penetration rate

CV_x = Coefficient of variation

Table 1-C

This table is from the U.S. Bureau of the Census - Current Population Survey

TABLE 1 C

PERCENTAGE OF HOUSEHOLDS WITH A TELEPHONE BY HOUSEHOLDER'S AGE

	ALL RACES		WHITE		BLACK		HISPANIC	
	Unit	Avail	Unit	Avail	Unit	Avail	Unit	Avail
8 MONTH AVERAGE								
TOTAL HOUSEHOLDS	91.7	93.8	93.3	95.0	80.4	84.8	81.1	84.[illegible]
16-24 YRS OLD	77.4	83.6	80.0	85.6	57.9	69.7	62.8	74.[illegible]
25-54 YRS OLD	91.8	93.8	93.5	95.1	80.2	84.5	82.7	85.4
55-59 YRS OLD	94.9	96.1	96.0	96.9	87.2	89.6	87.4	89.7
60-64 YRS OLD	95.1	96.0	96.0	96.8	87.9	89.8	88.8	90.4
65-69 YRS OLD	95.9	96.7	96.9	97.5	87.9	90.2	90.4	91.9
70-99 YRS OLD	95.5	96.6	96.1	97.2	89.1	91.3	86.8	90.1
NOVEMBER 83								
TOTAL HOUSEHOLDS	91.4	93.7	93.1	95.0	78.8	83.9	80.7	84.6
16-24 YRS OLD	76.6	84.1	80.2	86.2	49.9	68.2	64.9	71.9
25-54 YRS OLD	91.5	93.7	93.4	95.2	78.7	83.3	81.8	85.6
55-59 YRS OLD	95.0	96.1	96.1	97.0	86.3	88.5	89.3	89.3
60-64 YRS OLD	95.5	96.4	96.4	97.2	89.5	90.7	87.3	90.2
65-69 YRS OLD	95.5	96.2	96.5	97.0	87.2	89.0	90.7	90.7
70-99 YRS OLD	95.4	96.5	96.0	97.0	90.1	92.3	85.5	89.1
MARCH 84								
TOTAL HOUSEHOLDS	91.8	93.6	93.3	94.9	80.1	84.1	80.7	83.6
16-24 YRS OLD	77.8	84.0	80.3	85.5	57.9	71.5	59.0	66.2
25-54 YRS OLD	91.9	93.7	93.5	95.0	80.4	84.0	83.2	85.6
55-59 YRS OLD	94.9	95.9	95.7	96.6	87.6	89.9	88.7	90.5
60-64 YRS OLD	94.2	95.3	95.9	96.7	81.7	85.0	87.4	89.6
65-69 YRS OLD	96.1	96.6	97.0	97.4	87.8	89.3	85.8	87.8
70-99 YRS OLD	95.3	96.3	96.2	97.1	87.2	88.8	82.2	85.5
JULY 84								
TOTAL HOUSEHOLDS	91.6	93.8	93.2	95.0	80.5	85.3	81.1	84.6
16-24 YRS OLD	77.0	83.3	79.4	85.3	60.4	70.0	62.9	70.8
25-54 YRS OLD	91.7	93.8	93.4	95.1	79.8	84.9	83.1	85.8
55-59 YRS OLD	95.1	96.3	96.1	97.1	87.5	90.2	87.4	91.4
60-64 YRS OLD	95.0	96.2	95.8	96.9	87.7	89.5	88.1	90.5
65-69 YRS OLD	96.4	97.1	97.3	97.9	89.3	91.3	88.7	90.6
70-99 YRS OLD	95.2	96.5	95.9	96.9	89.6	93.1	84.0	88.5
NOVEMBER 84								
TOTAL HOUSEHOLDS	91.4	93.6	93.1	95.0	78.9	84.0	81.1	84.5
16-24 YRS OLD	76.1	83.4	79.0	85.4	56.3	70.8	60.8	70.8
25-54 YRS OLD	91.4	93.6	93.3	95.1	78.5	83.3	83.1	85.8
55-59 YRS OLD	94.9	96.2	96.3	97.5	84.7	87.4	85.3	88.3
60-64 YRS OLD	95.6	96.5	96.5	97.3	90.3	92.1	86.0	87.2
65-69 YRS OLD	96.0	96.7	97.1	97.6	86.7	89.1	96.2	96.2
70-99 YRS OLD	95.3	96.6	96.1	97.2	88.0	90.7	87.1	88.8
MARCH 85								
TOTAL HOUSEHOLDS	91.8	93.7	93.3	95.0	80.1	84.4	81.2	84.1
16-24 YRS OLD	77.3	83.1	79.6	84.8	59.8	70.0	62.4	67.1
25-54 YRS OLD	91.9	93.8	93.6	95.2	79.5	83.9	83.0	85.5
55-59 YRS OLD	94.9	95.9	95.8	96.7	87.3	89.1	86.5	89.1

Table 2

% OF HOUSEHOLDS WITH TELEPHONES
IN THE C&P SERVED AREA OF VIRGINIA

Year (1)	Residence Access Lines-In Service (June 30) (2)	Bell Served Households (July 1) (3)	Telephone Penetration (2)/(3)X100 (4)
1960	564,824	763,000	74.0
1961	588,783	782,000	75.3
1962	616,381	802,000	76.9
1963	640,932	823,000	77.9
1964	669,764	843,000	79.5
1965	702,113	863,000	81.4
1966	736,798	882,000	83.5
1967	776,020	904,000	85.8
1968	819,205	933,000	87.8
1969	860,518	975,000	88.3
1970	896,711	1,011,000	88.7
1971	933,814	1,027,000	90.9
1972	972,268	1,068,000	91.0
1973	1,025,160	1,104,000	92.9
1974	1,069,475	1,134,000	94.3
1975	1,097,747	1,168,000	94.0
1976	1,132,060	1,202,000	94.2
1977	1,166,408	1,225,000	95.2
1978	1,212,158	1,263,000	96.0
1979	1,255,582	1,305,000	96.2
1980	1,288,870	1,355,000	95.1
1981	1,319,998	1,386,000	95.2
1982	1,343,961	1,416,000	94.9

Source: (1) Data for Residence Access Lines-In-Service are taken from the Virginia monthly No.7 report, years 1960-80, and taken from the Virginia No. 2705 report for years 1981-82;

(2) Household data is based on household counts from the Decennial Census 1960, 1970 and 1980; and are based on the Current Population Survey and P-25 Population Reports in other years. These reports originate at the U.S. Bureau of the Census.

October 1983

Table 3

Telephone Penetration by Age and Income, Virginia, 1980

	16 to 24	25 to 29	30 to 34	35 to 39	40 to 44	45 to 49	50 to 54	55 to 59	60 to 64	65 to 69	70 to 74	75 to 79
	HSHLDER	HSHLDER	HSHLDER	HSHLDER	HSHLDER	HSHLDER	HSHLDER	HSHLDER	HSHLDER	HSHLDER	HSHLDER	HSHLDER
Household Income												
Less Than $10,000	0.68	0.75	0.77	0.79	0.78	0.76	0.80	0.83	0.86	0.89	0.92	0.91
$10,000 - 14999	0.79	0.87	0.86	0.86	0.89	0.90	0.90	0.94.	0.95	0.97	0.97	0.97
$15000 - 19999	0.83	0.92	0.94	0.94	0.94	0.94	0.95	0.97	0.96	0.97	0.96	0.95
$20000 - 24999	0.92	0.96	0.97	0.97	0.96	0.97	0.98	0.98	0.98	0.97	0.99	0.99
$25000 - 29999	0.95	0.98	0.98	0.99	0.98	0.99	0.97	0.99	0.99	0.98	0.99	0.98
30000 - 34999	0.91	0.98	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.99	0.98
$35000 - 39999	1.00	0.98	0.98	1.00	0.98	0.99	0.99	0.99	1.00	0.99	1.00	0.97
$40000 - 44999	0.94	0.97	1.00	0.99	1.00	1.00	0.99	0.98	0.99	1.00	1.00	1.00
$45000 - 49999	0.95	0.99	0.99	0.99	0.99	0.99	1.00	1.00	0.98	1.00	0.97	1.00
$50000 or More	0.98	0.98	1.00	0.99	0.99	0.99	0.99	1.00	0.99	0.98	0.99	0.97
Total	0.78	0.89	0.92	0.94	0.94	0.94	0.94	0.95	0.94	0.94	0.93	0.93

These data were generated from the 1980 Public Use Microdata Sample by the Pennsylvania State Data Center, at the request of A.D. Schmidley, Bell Atlantic

Table 4

TELEPHONE PENETRATION
BY AGE OF HOUSEHOLDER (%)
(1980 CENSUS)

AGE	RATE	AGE	RATE
18-24	83.8	50-54	96.0
25-29	92.0	55-59	96.2
30-34	94.5	60-64	96.2
35-39	95.3	65-69	96.0
40-44	95.5	70-74	96.2
45-49	95.8	75+	95.6

This data was generated from the Public Use Microdata Samples PUMS for the Bell Atlantic served jurisdictions of Washington, D.C., Delaware, Maryland, New Jersy, Pennsylvania, Virginia and West Virginia, combined.

Figure 1

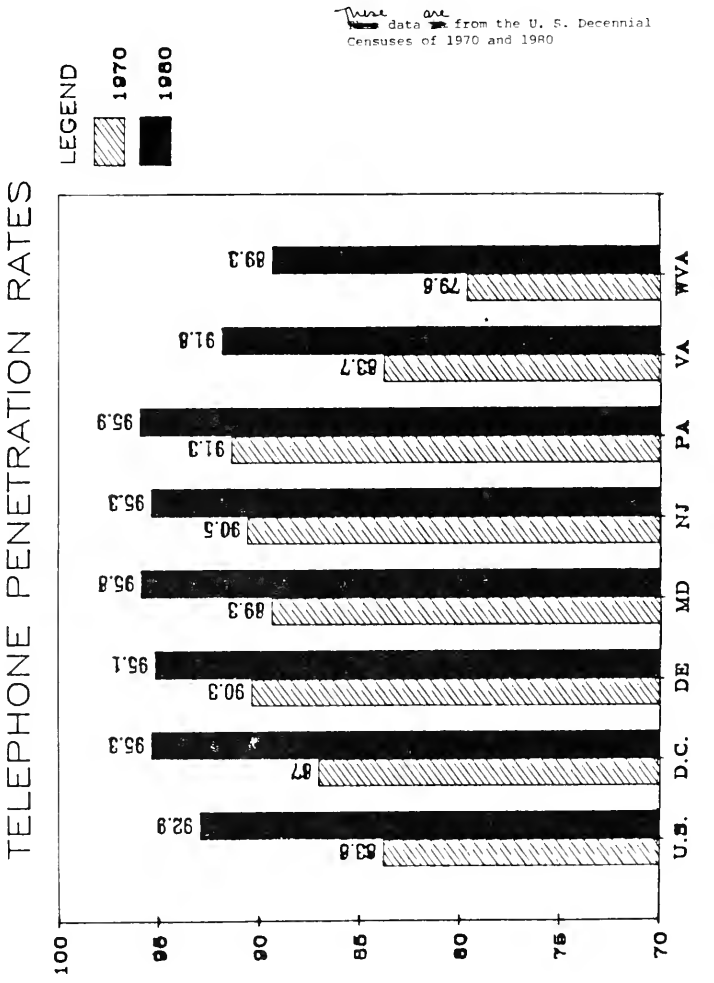

Data are from the U. S. Decennial Censuses of 1970 and 1980
LEGEND
1970
1980
TELEPHONE PENETRATION RATES
100
95
90
85
80
75
70
U.S.
63.8
92.9
D.C.
87
95.3
DE
90.3
95.1
MD
89.3
95.8
NJ
90.5
95.3
PA
91.3
95.9
VA
83.7
91.8
WVA
79.8
89.3

Figure 2

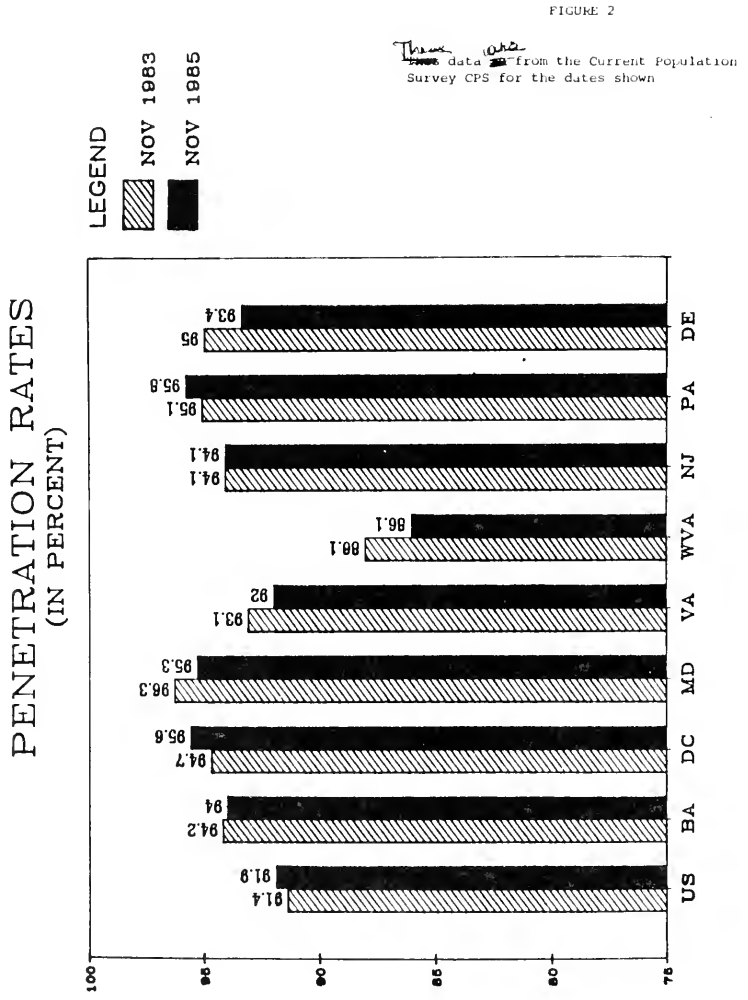
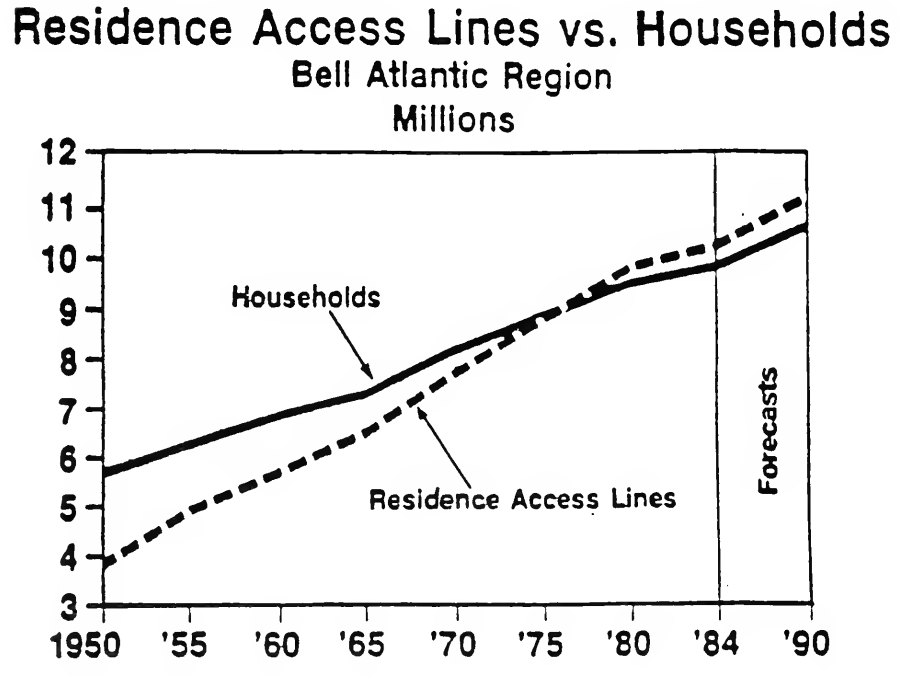


FIGURE 2
Thes data are from the Current Population Survey CPS for the dates shown
LEGEND
NOV 1983
NOV 1985
PENETRATION RATES
(IN PERCENT)
US 91.4 91.9
BA 94.2 94
DC 94.7 95.6
MD 96.3 95.3
VA 93.1 92
WVA 86.1 86.1
NJ 94.1 94.1
PA 95.1 95.8
DE 95 93.4
100
95
90
85
80
75

Figure 2


Residence Access Lines vs. Households
Bell Atlantic Region
Millions
12
11
10
9
8
7
6
5
4
3
Households
Residence Access Lines
Forecasts
1950
'55
'60
'65
'70
'75
'80
'84
'90

Evaluation criteria for the selection of computer mapping systems

by Marci L. Belcher[1]
Marketing Manager
MPSI Systems, Inc.
650 California Street, 31st Floor
San Francisco, CA 94108
(415) 434-3222

Significant differences exist between the computer mapping products available in today's market. The way in which data are managed and manipulated by a system, in which information is represented in the mapped end product, the resources required for use, and the level of service provided by the vendor, are all important criteria that distinguish one product from another. Ultimately, the computer-mapping system selected should satisfy the application needs and environment of an organization. However, identifying the decision-making criteria to use in the selection process can be difficult unless one is aware of features available across products, and understands how these translate into potential benefits or problems.

Computer mapping systems use data as raw material in order to present relationships in a visual, geographic form. Therefore, a mapping system designed to truly support market analysis should be capable of organizing and managing all the information required for decision-making. One of the first issues to address in selecting a mapping product then is whether a mainframe computer or microcomputer-based system is required.

Centralized organizations with large varied data files on customers, demographics, lifestyle profiles, competitive informaton, consumption patterns, performance, etc., should consider a mainframe based product. Mainframe based systems are designed to process large data files efficiently, execute tasks quickly, and perform more complex functions. The cost of mainframe based products are higher than micro based. However, most of these systems help organizations optimize their investment in data by bringing the information to multiple users for different applications. Mainframe computers also tend to provide more powerful analytical capabilities, increasing the utility of maps in complex problem-solving.

[1]Presented at the IASSIST 86 Annual Conference and Workshops, May 23, 1986

Decentralized or smaller organizations may find that microcomputer based mapping systems offer some of the capabilities of mainframe packages at a lower cost. Each micro workstation is an autonomous unit in which the data base may be designed and enhanced on an individual basis. Packages that interface easily with spreadsheet or reporting programs allow users to build a data base for mapping that includes information from a variety of sources. The greatest issue related to selection of a micro based system is the limit to the amount of data that can be stored and used at one time. New laser disk storage has provided users with access to a large volume of vendor supplied data. However, these data cannot be integrated with other information so that relationships, for example, between a company's retail performance and the corresponding demographic profile of that sales area, can be viewed on the same map.

Mapping systems index data to geographic units. Products differ greatly, however, in the number of geographic types to which information may be linked and, therefore, the types of data which may be represented in maps. All mapping systems display information through areas such as census tracts or postal zones. Mapping systems that also index data to locations and networks provide users with a higher level of analytical and interpretive capability. A retail location's performance, product mix or pricing may be thematically represented at different outlets. Traffic counts, type of roadway or directional information may be shown by varied transportation networks. A system that represents information linked to two or three geographic layers in the same map allows users to evaluate performance at a retail outlet in the context of the underlying market demand, the success and spatial location of competitors, and the level of traffic flow near the facility.

The structure of a system's data base can limit the flexibility the which the information may be used. In systems with parallel or hierarchical structures, information requests or queries must conform to the way in which data are organized in the data base. With a relational structure, any data item, be it a map element or an attribute in a related table, serves as an entry point or "front door" to the entire data base. Every attribute related to an individual geographic unit is also linked to all others. As a result, questions may be based on decision criteria or rules, in a way that replicates the human thought process.

The benefit of a relational data base structure is that it may be possible, if the software product is appropriately designed, to query the data base using multiple criteria related to the attributes of geographic areas. For example, the user may want to see only census tracts that have median home values greater than $100,000 in which he has less than a 15% customer penetration level. A solution set of only the units meeting these criteria is formed.

These units may be further color segmented by another variable such as percentage of population between the ages of 20 and 34 years of age. Geographic units that meet a first-order set of criteria are first identified. These units may then be automatically segmented by variables that support more concrete decision-making. This type of system performs the work that multiple overlays of single variable maps ultimately achieve, but with greater efficiency and speed.

A major problem in data analysis has been the inability to integrate data that did not correspond to common units of geography. Often data important for planning are excluded because they are not comparable to other information in a common framework. This problem is solved with a computer mapping system that integrates data through its inherent common element - geography. Disparate geographic units are merged through an overlay process in the system, producing solutions to queries that are made up of the unique

geographic segments that meet all imposed criteria. Resulting geographic units may then transcend standard geography and consist of the intersections of units such as zip codes and census tracts.

Map overlay eliminates the need to continually geocode data to a consistent geographic unit. Customer information linked to zip codes may be analyzed directly with demographic information based on census tracts. In addition, if the overlay process includes two or three types of geography (areas, networks, locations), relationships between information based on disparate geographic types may be studied. For example, only retail locations with poor revenues that are also in areas with high potential demand for the products or services offered could be mapped. The time consumed in map interpretation is greatly reduced because a key problem or situation is directly identified through the query. In maps with every area and location represented, a labor intensive and often non-methodical search process must be conducted to identify relevant areas.

Often data are not stored in the form most useful for analysis. Some products feature the ability to perform mathematical operations using attribute items from the data base in any valid arithmetic expression. New data items formed may be used immediately in a query without making permanent modifications to the data base. For example, raw counts of quarterly sales performance can be changed to percent growth. Demand estimates could be developed by applying an equation such as (target population X percent expected conversion X average dollar sale).

As an aid in selecting meaningful data categories for segmentation, some systems display the distribution and range of values for each variable. Products may also automatically create categories with even counts of data, or let users select breakpoints, and automatically didplay the resulting histogram. Histogram displays may represent the distribution of a variable on an individual or cumulative basis by the number of geographic units or by the size of the area in each category.

Representing raw counts or even percentages of data in maps often leads to the misinterpretation of information. Systems that perform density calculations help users understand that even though a large zip code region has a large population, that population may be much less concentrated than in a small region with a smaller number of residents. Through dot-density mapping, the concentration of data may be viewed on a macro-level.

The ability of a system to zoom-in and increase the resolution of map elements before plotting is important in applications which include the study of locations or street networks that may reside in close proximity. Retail bank branches are often located on the four corners of one intersection. Displaying all bank offices in one county would mean that many locations would be plotted on top of one another. The ability to increase the resolution should be contrasted with what many systems accomplish when they enlarge a screen display. It is the difference between blowing up a photograph, to enlarge elements, and using a zoom-lens so that greater detail is actually captured.

Managers under pressure to make fast decisions, or overloaded with work, can quickly test hypotheses in systems with on-line interactive planning capabilities. Computer screen geographic solutions are immediately produced and may be scanned for important relationships. An entire nation may be analyzed using broad criteria, then each area of interest seen in greater detail using stricter criteria on smaller geographic units. A systematic macro to micro analysitic framework helps pinpoint target areas for further study in a highly efficient manner.

The extent to which maps are used to understand spatial relationships and distance will

be an important criterion in deciding whether the system required should produce cartographically accurate maps. Many micro based products today create a graphic image of map elements but do not relate these elements to a projection system or scale. Cartographically accurate mapping systems usually display longitude and latitude "tick" marks and present a scale for use in calculating exact distances between specific loctaions.

Mapping systems that summarize information in rings around a location, within a bank of geography, by solutions sets, or in various other shapes give insight into the specific composition of target markets. Systems that integrate information from various data sources are able to produce reports on a full spectrum of information about areas. Some reporting functions enable users to specify custom layouts, others rely on default parameters. Packages with a high degree of data manipulation and calculation capability eliminate the need to transfer data for this purpose to other software packages.

Another feature to consider in computer mapping systems is the ability to display data items in the data base near the corresponding geographic elements. These items are not necessarily part of a query, but provide additional detail about specific areas, locations, or networks. For example, the average family size and lifestyle profile label could be printed in areas that are also being color-segmented by sales performance.

Computer mapping systems differ greatly in the amount of attention that has been paid to information display on finished maps. Hatch and cross-hatch patterns are available for plotting in a variety of angles. Multiple symbols and patterns such as dashes and double lines may be used to segment locations and road networks. Text for title and legend boxes may be automatically developed by the software and displayed in a selection of font styles and sizes anywhere surrounding the map image.

To the extent that it is possible, the colors and patterns selected for terminal displays should extend directly to map specifications so that this information does not have to be re-entered. Systems with a direct interface to off-the-screen color printers exactly replicate a screen image on paper. This method of map production is the easiest and fastest, displaying first on the computer screen exactly what will be produced in the printed map.

The degree of user friendliness may mean the difference between whether the computer mapping system becomes a successful and often used decision support tool. Products should include on-line help programs, fully prompted input screens, and English-like syntax. Familiarity with a programming language should not be required. It should be possible to save and recall for editing queries and plotting specifications. Finally, a system should be ergonomically easy to use by individuals with a wide range of functions and skill levels.

Most companies offering computer mapping packages also supply geographic files. Vendors may specialize in providing highly detailed map files that include features such as cities, parks, airports, lakes, rivers, bridges, rail and rapid transit lines as well as multi-layered transportation networks. These features add a significant level of detail that reduce or eliminate the need to use base maps for further reference.

A product which includes the capability of digitizing and encoding files will allow the user to add or modify map elements on an ongoing basis. This may be a particularly desirable feature if,for example, the market typically under study has rapidly changing elements such as retail locations, or if an organization's sales regions or other administrative units frequently change.

The ease with which the tabular data and map files in a system may be modified is important if the system is to be continually enhanced. Information such as sales performance may be a desired addition on a quarterly basis. It should also be possible to easily add additional geographic areas and units to the system, with programs linking new data to the corresponding map files. Learn whether a major reorganization of sales regions, or additions in product line, would require restructuring of the data base, and how easily this is accomplished.

Turnkey systems, which integrate software and hardware into a readily usable package, are convenient but rarely allow custom modifications that can make a system much more valuable to a specific organization. For example, a system that cannot integrate proprietary company information or data from other vendors offers limited scope.

Include, in the cost of a system, the cost of other software required for operation either on a micro or mainframe computer. Additional hardware may also be required such as graphics boards, emulator cards, terminal controller, modem, and, of course, plotters and/or printers. For mainframe products, it will be important to understand memory (region) and disk storage requirements.

Selecting a vendor who bases his success on the satisfaction of his customers is an important ingredient in selecting the right computer-mapping product. Vendors with experienced personnel in your industry or function may provide valuable application support. Vendors that market demographic data, map files and mapping software offer one-stop shopping that can reduce the time spent with several different vendors. Software maintenance programs provide enhanced versions of a product, thereby maintaining capabilities close to state of the art. And it is most important that vendors should provide training, installation, testing, tutorials, system documentation, and toll-free telephone assistance.

Vendors confident the functionality and benefits of their product may offer test or pilot projects to enable companies to gain hands-on experience at little or no cost. Taking advantage of these programs will promote confidence that the system selected is the right one for the application needs and structure of your organization. □

Ensuring racial representation on jury panels:
an empirical and simulation analysis

Hiroshi Fukurai[1]
University of California, Riverside

Edgar W. Butler[1]
University of California, Riverside

Jo-Ellan Huebner-Dimitrius[1]
California State University, Long Beach

California law specifically states that persons listed for service in the court:

> shall be fairly representative of the population in the area served by the court and shall be selected upon a random basis (Section 9, 203).

Introduction

Geographic groupings which overlap with racial and economic groupings constitute recognizable classes.[2] Rural residents, for example, might be underrepresented due to excuses based on distance to the courthouse; often selection officials acquiesce in the reluctance of rural residents to serve. Excuses based on willingness to travel great distances have so reduced the jury pool that remedial action is required even without proof of geographic cohesiveness.[3]

The notion of vicinage or geographical locality requirement of jury selection has been traced at least as far back as to Charlemagne (Charles the Great) in 768 A.D. He instituted several reforms, one of which was the establishment of "Inquisito". One of the requirements of the Inquisito was that 13 to 66 witnesses be chosen from the neighborhood where they would have knowledge of the matter in dispute (Moore

[1]Paper prepared for presentation at the International Association for Social Science Information Service and Technology (IASSIST) Conference held in Marina Del Rey, California, May 23, 1986.

[2]See Thiel v. Southern Pacific Co., 328 U.S. 217 (1946), State v. Holstrom, 43 Wis. 465, 168 N.W. 2d 574 (1969), and State v. Cage, 337 So. 2d 1123 (La. 1976). The Federal Act requires that selection procedures "ensure that each county, parish or similar political subdivision within the district or division is substantially proportionally represented in the master jury wheel for that judicial district, division, or combination of divisions" (U.S. 1968, Section 1863 (b) (3)).

[3] See United States v. Fernandez, 480 F. 2d 726, 732-33 (2d Cir. 1973).

1973). The requirement of a jury from the vicinage, found in the Magna Carta as well as the Sixth Amendment, is also based on the notion that jurors should be selected from local residents. The meaning of this requirement is often unclear, as when a case from one division in a federal district court is tried in another division, or when grand jurors are selected from only one division.[4]

Currently, federal law determines the nature of prospective jurors by specifying two key concepts in jury venire or panel selection procedures: (1) "a random" selection of jurors, and (2) the inclusion of special geographic districts wherein a particular court convenes, i.e., vicinage requirements (U.S. 1968, Section 1861). In California, as in most states, the law similarly requires: (1) a random selection of jurors and (2) selection from "judicial districts of the respective counties" (CA. 1981, Section 197, 206). Recent Federal and California Supreme Court decisions are such that any substantial violation of these basic requirements of jury selection in representativeness is a prima facie case of discrimination.[5] Subsequently, an increasing number of challenges concerning the underrepresentation of "cognizable groups," e.g., minorities, have been brought claiming violation of the Sixth Amendment, a representative jury selected from a fair cross section of the community.[6]

One of the major problems in jury challenges is to establish a prima facie case for the underrepresentation of minorities. Part of the problem is due to the ambiguous relationship between random selection and vicinage requirement (area or district). Past Supreme Court cases have dealt with the systematic underrepresnetation of cognizable groups, e.g., blacks and Hispanics; however, the Court has not addressed the extent to which the area served by the court relates to the random selection of potential jurors. Vicinage or geographic representativeness has rather been dealt with, along with the random selection of jurors, without geographic representativeness being clearly demarcated.

In Duren v. Missouri, for example, the U.S. Supreme Court held that a three-prong test must be applied to establish a prima facie case of discrimination: (1) the group alleged to be excluded is a 'distinctive' group in the community, (2) the representation of this group in venires from which jurors are selected is not fair and reasonable in relation to the number of such persons in the community, and (3) this underrepresentation is due to systematic exclusion of the group in the jury-selection process (Duren v. Missouri 439 U.S. 357 364 1978). However, the Court did not spell out a clear cut relationship between juror representativeness and the vicinage requirement, e.g., what is the "community"?

[4] See U.S. 1968, Section 1861 and House Report at 1801.

[5] See U.S. 90th Congress Senate Report No. 891 1967; U.S. 90th Congress House Report No. 1076 1968; The Yale Law Journal 1970; Kairys 1972; De Cani 1974; Chevigny 1975; Alker, Hosticka, and Michell 1976; Kairys, Kadane, and Lehoczky 1977; Alker and Barnard 1978; Heyns 1979; Butler 1980a, 1980b, and 1981; Butler and Fukurai 1984; Fukurai and Butler 1985.

[6] For example in California, see People v. White 43 Cal. 3d 740 1954; People v. King 49 Cal. Rptr. 562 1966; People v. Sirhan 7 Cal. 3d 258 1978; People v. Wheeler 148 Cal. Rptr. 890 1978; People v. Estrada 155 Cal. [6](cont'd) Rptr. 731 1979; People v. Graham 160 Cal. Rptr. 10 1979; People v. Harris 36 Cal. 3d 36, 201 Cal. Fprt. 782 679 P. 2d 433 1984. In Federal Supreme Court, see Alexander v. Louisiana 405 U.S. 625 1972; Peters v. Kiff 407 U.S. 493 1972; Taylor v. Louisiana 419 U.S. 522 1975; Duren v. Missouri 439 U.S. 357 1979; City of Mobile, Ala v. Bolden 466 U.S. 55 1980.

Vicinage Requirement

The vicinage or geographical requirement of jury trials is an essential element of the Sixth Amendment as it pertains to the jury selection process. For illustrative purposes, we will use Los Angeles County and its twenty mile radius rule. However, the process itself is generalizable to all areas of Los Angeles County and any other bounded area such as a county, judicial district, etc.

The first provision for a jury trial in a vicinage can be found in Article III of the Constitution. Article III. Section 2 notes:

> The Trial of all Crimes, except in Cases of Impeachment, shall be by Jury; and such Trial shall be held in the State where the said Crimes shall have been committed; but when not committed within any State, the Trial shall be at such Place or Places as the Congress may by Law have directed.

Early in the 1970s, the Los Angeles County Board of Supervisors developed a policy that no juror had to travel more than twenty miles from his/her house to the courthouse. The County Board adopted this rule because of convenience for prospective jurors and economic reasons for the county. Subsequently, in California, the legislature defined the judicial district in Los Angeles County as being within a twenty mile radius from each courthouse. The California Code of Civil Procedure states that:

> Each court shall adopt rules supplementary to such rules as may be adopted by the Judicial Council, governing the selection of persons to be listed as available for service as trial jurors. The persons so listed shall be fairly representative of the population in the area served by the court, and shall be selected upon a random basis. Such rules shall govern the duties of the court and its attaches in the production and use of the juror lists. In counties with more than one court location, the rules shall reasonably minimize the the distance traveled by jurors. In addition, in the County of Los Angeles no juror shall be required to serve at a distance greater than 20 miles from his or her residence (CA. 1981, Section 7. 203).

Despite the explicit rule of random selection of potential jurors from the judicial district defined within 20 mile radius, recent jury venire challenge cases have argued the following two points: (1) there is a significant underrepresentation of prospective minority jurors and (2) there is an overrepresentation of particular neighborhoods with high concentrations of anglos (Heyns 1979; Butler 1980a, 1980b, 1981; Butler and Fukurai 1984; Huebner-Dimitrius 1984; Fukurai and Butler 1985; Fukurai and Butler 1986). These studies have shown that census tracts with a high anglo concentration are consistently overrepresented and consequently jury venires have consisted of a large number of potential anglo jurors, and an underrepresentation of minorities. This is apparent for all Superior Court districts in Los Angeles County except the Central District (Heyns 1979).

It is theoretically possible to have race/ethnic representation on juries, yet not have a fair cross section of the community or areas served by the court from which jurors are being drawn to serve on juries (Heyns 1979; Huebner-Dimitrius 1984; Fukurai 1985; Fukurai and Butler 1985). Generally however, racial and geographic representativeness are highly correlated; therefore, it is possible to ensure the cross-section representation of minorities by controlling the random selection of geographic areas. Currently, the overrepresentation of particular neighborhoods contributes to a substantially greater chance of anglos serving on

juries, while a random selection of neighborhoods would ensure the fair representation of minorities as prospective jurors within a district.

In this paper, we present an analytic strategy that will overcome racially disproportionate jury venires. Rather than first focusing on the selection of individual potential jurors, random selection of neighborhoods is examined, i.e., census tracts from which prospective jurors are being drawn to serve on juries. Our analysis demonstrates the extent to which neighborhood representativeness could rectify the disproportionate underrepresentation of minorities currently the case in most jury venires. The main thrust of this paper, thus, is threefold: (1) to propose a geographic sampling strategy to overcome underrepresentativeness of minorities, (2) to illustrate our strategy using simulation techniques, and (3) to show the extent to which geographical randomness can help ensure that racially proportionate jury venires are obtained. By simulating the Los Angeles County selection process, a comparison between the actual jury composition and the simulated jury composition is examined to show the extent to which the proposed geographic sampling strategy is superior to the current selection procedures employed in Los Angeles County and elsewhere.

Data

Two data sets were linked to serve as the foundation for the simulation of the jury selection process: (1) 1980 U.S. Census Bureau data and (2) jury impanelment lists for a retrial of the Harris case (36 Cal. 3d 36 201 Cal. Rptr. 782 679 P. 2d 433 1984).[*]

Eight jury impanelment lists were obtained to delineate neighborhoods (census tracts) from which jurors were being drawn to the Long Beach Superior Court and to determine whether or not the panels represented a fair cross section. The impanelment period under investigation, while not ideal, was lengthy enough to determine whether or not jury venires were representative of the community population. These eight panels were typical of panel data available for other time periods, including the first Harris trial 1979.

Empirical Analysis

Figure 1 depicts racial composition of the Long Beach judicial district using a variety of definitions of the area served by the court. Map A illustrates the areas served by the Superior Court, as presented in the Harris retrial. *(Ed.Note. Figures and maps have collected together at end of article)*

This variation in the definitions of the area served by the court shows that there is a potential for either conscious or unconscious manipulation of minority representation on jury panels. Thus the particular "area served by the court" becomes important in jury challenges. That is, if it is to be determined whether or not jurors represent a fair cross section of the community, the area served by the court must be clearly delineated or a valid comparison cannot be made.

Six different areas served by the court emerged during the Harris retrial. The first was Los Angeles County as a whole. In the first Harris

[*] Empirical analyses of People v. Harris (36 Cal. 3d 36, 201 Cal. Rptr. 782 679 P. 2d 433 1984) were performed at University of

[*](cont'd) California, Riverside. In People v. Harris, the motion of respondent for leave to proceed in forma pauperis was granted; however, the Writ of Certiorari by the prosecution to the Federal Supreme Court was denied on Oct. 29, 1984.

trial the prosecution argued that Los Angeles County-wide data were the proper comparison. The Harris opinion rendered by the California Supreme Court concluded that

> The parties, however, presented evidence and argued this case on the assumption that all juries in Los Angeles County must be representative of the entire county. The principal question before us is whether evidence based on total countywide population figures, rather than jury-eligible population, is adequate to make out a prima facie case; for the reasons explained in this pinion, we conclude that it is. The state has not attempted to rebut this prima facie showing by arguing that the Long Beach juries need only represent those persons living within 20 miles of the courthouse, and has not attempted to show that such juries were truly representative of that limited area (Harris 36 Cal. 3d 36. 201 Cal. Rptr. 782 679 P. 2d 433 1984).

A second area served by the court is within a 20-mile region, as delineated by California state law; that is, any juror may be excused from being sent to a particular courthouse that is further than 20 miles from his/her residence. In 1978, the 20-mile region delineated by the Jury Services Division in Los Angeles County was for the most part a 20-mile straight line from the courthouse. However, in 1983 the area served by the court was reduced to a 15-mile direct line, presumably on the basis that the driving distance would be 20 miles. Thus, a third definition of the area serviced by the court was considered.

A fourth area served by the court was empirically delineated. This area was determined by delineating those census tracts from which jurors were summoned for eight panels. A fifth area served by the court could not be determined geographically but is obviously different from the others. This fifth area is a subset of the fourth area which was geographically determined. For each juror summoned, knowledge of their census tracts and address was made available, thus the area served by the court could be empirically determined. All of the potential jurors who showed up at the Long Beach Courthouse came from these impanelments and thus were a subset of the impanelment lists. However, between the impanelment or summons stage and the jury venire stage, there was between a 40-50 percent dropout. Thus, they are similar but not the same. Unfortunately we were unable to delineate areas of residence at this stage.

Finally, during the course of the Harris retrial, the prosecution argued that a sixth area was more important than these other five. This specific area was known as the Long Beach Superior Court District, as defined by the Los Angeles Board of Supervisors. This area is used by the legal system in allocating trials. Thus, if a person commits a crime in this bounded area and it becomes necessary to have a trial, it typically, but not invariably, will be assigned to the Long Beach Courthouse. However, this particular area is not coterminous with any of the other five areas served by the court. Obviously, all the areas are within Los Angeles County, but otherwise they have nothing in common.

Table 1 shows the racial composition of the eligible Hispanic population and impanelment lists using the 15-mile radius definition of the area served by the court. Thus, while 20.9 percent of potential jurors at a 15-mile level were Hispanic, only 10.1 percent of jurors impaneled and summoned to the Long Beach Superior Court were Hispanic. Underrepresentation of Hispanic jurors was inevitable because of the under-selection of Hispanically dominant census tracts. Table 1 thus shows that more than one-half of the potential Hispanic jurors were underrepresented on the impanelment list. Z scores and

chi-square values show that Hispanic jurors are statistically underrepresented on the impanelment list; thus, the Hispanic composition on the impanelment lists is significantly different from the racial composition of the Long Beach judicial district, as defined by the 15-mile radius.

The underrepresentation of both Hispanic and black jurors on the eight panels under investigation was consistent. Table 2 shows the racial composition of minority jurors on both impanelment lists and census tracts from which potential jurors are summoned. Census tracts with a high concentration of anglos are overrepresented whereas minority dominated census tracts are underrepresented.

Table 3 shows the representation of census tracts on eight impanelments. Potential jurors from one census tract were represented thirty-six times, while fifty-one census tracts were represented less than five times. Table 3 also indicates that one-half the potential jurors came from twenty-three census tracts (5.2%) of the total of 439 tracts in the Long Beach judicial district defined by a 15-mile radius.

Table 4 indicates the average representation of census tracts on eight panels. The table shows the extent to which tract representation is related to the racial composition of the census tract. For example, census tracts which were selected less than the average number of times had three and ten percent higher black and Hispanic populations respectively. One-half the overrepresented census tracts had four and fourteen percent less black and Hispanic population, respectively.

Maps 1 to 8 illustrate the census tracts from which actual potential jurors were summoned. These maps show that the census tracts were concentrated in particular regions, i.e., the lower portions which border on Orange County. From previous tables, it should be obvious by now that these census tracts are characterized by a high concentration of anglos. In any case, and for whatever reason, anglo dominant census tracts are clearly overrepresented on the jury impanelment lists.

One dubious explanation is that anglos are more qualified than minority groups for jury duty.[8] However, the proportion of qualified jurors is the same for the impaneled census tracts and the Long Beach Superior Court judicial district as a whole. Table 5 shows the proportion of qualified jurors in the impanelment list and the Long Beach judicial district. While 19.9 percent of jurors in impaneled census tracts are qualified jurors, 19.1 percent of those in the Long Beach judicial district are equally qualified. Thus, the percentage of qualified jurors has no bearing on the underrepresentation of minority jurors. Further, a random method of selecting jurors has not been exercised, i.e., impaneled census tracts are clustered in particular regions characterized by an anglo population. Many minority dominant census tracts are not included in the impanelment list, even though the proportion of qualified jurors is the same in both impaneled and non-impaneled census tracts.

[8] Research indicates the overrepresentation of anglo jurors is necessary since criminality is inherent in some minority groups (Hepburn 1978; Cullen and Link 1980; Turk 1981; Kramer 1982). Thus, minority groups "take[s] a permissive view of crime within its border. As a result, the black community is vulnerable to its own criminal element as well as to the criminal element of the white community" (The Yale Law Journal 1970, p.534). Further, researchers suggest that county clerks responsible for selecting names from master files purposely exercise systematic selection rather than random selection in creating racially disproportionate jury pools (Alker and Barnard 1978; Levine and Schweber-Koven 1976). Because so many different persons use individual discretion to decide who should be excused and who should serve, the possibility of individual prejudice influencing excuses and exemptions is great (Van Dyke 1977, p.391).

Geographic Random Selection

One means by which to rectify the disproportionate representation of census tracts is to implement the random selection of census tracts within a judicial district, but geographically defined. Such random selection should provide a list of census tracts equally distributed within the limited, spatially bounded context, i.e., 15-mile radius, or whatever. Since our analysis shows that qualification of particular racial populations does not have a bearing on the selection of anglo-dominant census tracts, the random selection of tracts provides a foundation for equally selecting various racial/ethnic groups within them thus resulting in a fair cross-section of the population, vis-a-vis minorities.

A simulated random selection of census tracts was carried out in the following manner. Each census tract within a 20-mile radius of Long Beach judicial district was given a unique number. A series of random numbers were generated for the selected number of census tracts for each of eight panels. Those eight individual simulations were conducted to correspond to the actual eight impanelments as previously empirically analyzed and used in the Harris retrial.

Maps 9 through 16 illustrate the simulated mapping of census tracts randomly selected within the Long Beach judicial district. Each map shows that selected tracts are evenly distributed in space. The number of potential jurors also shows that using this process, minority groups would have an equal chance of selection for jury service.

Table 6 shows the racial composition of selected census tracts for each of the simulated panels. Within the 20-mile radius 26.7 percent of potential jurors were Hispanic and 14.8 percent for black. A Z scores statistical test for differences in racial composition between census tracts derived by random selection and the 20-mile radius district was then carried out. Not one of the scores was significant, suggesting that each of the randomly selected samples of tracts had a racial composition similar to that of the 20-mile radius judicial district. This, of course, is in stark contrast to the actual impanelments analyzed in the first section of this paper.

Map 17 shows the mapping of all census tracts in eight panels using the simulation method. The map shows that random selection of census tracts provides a virtual equally distributed list of tracts from which potential jurors would have been summoned. Such random selection also provides an unbiased racial representation.

Critique

The results of our simulation clearly show that the process we have suggested is far superior to the current process in ensuring a fair cross section of jurors.

One question, of course, is whether or not this process is allowable under current Federal and State statues. Our response is that not only is it allowable, but the results of the simulation imply that our process should be mandated by law. Another argument that possibly could be made against the proposed process is that qualification varies by district. However, our evaluation of the actual juror qualification rate for Los Angeles County compared with the Long Beach District (20-mile) showed that the qualification rate was virtually identical. Even if there had been some variation, such variation could be fitted into the system.

Another possible objection to the randomized geographical process is that it would increase

the overall mileage driven by jurors. This is true. Any system that results in a fair cross section will result in more aggregate miles driven because of the very fact that the jurors would be from all areas of the district rather than concentrated in certain areas. This is a necessary part of a system that results in a fair cross section of the community — jurors must come from all parts of the community. The proposed system does away with the idea of selecting only jurors from areas closest to the court, and in fact, requires just the opposite. That is, jurors are drawn from all areas of the district. However, the district could still fall within the state law mandated 20-mile region for Los Angeles County.

In Los Angeles County, a particular problem that must also be dealt with is the overlapping of judicial district boundaries. This problem is amenable to statistical sampling methods. However, even if some overrepresentation should occur, it would be substantially less than is now occurring using non-random selection of areas.

Finally, the analysis presented here represents only part of a year and thus might be considered static. A dynamic jury selection process involves selecting jurors periodically. However, jurors also are qualified only periodically. Thus a dynamic system of jury qualification could use the same technique described in the simulation section. That is, the jury qualification process could also be accomplished by the random selection of census tracts, and the mailing out of questionnaires periodically throughout the year.

Conclusions

Our analysis conclusively shows that currently there is a systematic and biased selection method employed in the Long Beach judicial district and elsewhere in Los Angeles County. The racial composition of actual impaneled census tracts indicates that (1) selected census tracts are clustered in regions with high concentrations of anglos and (2) Hispanic and black potential jurors are systematically weeded out in the selection process because of biased impanelment lists. One possible reason for such systematic selection of anglo dominant census tracts might be that anglo jurors in particular census tracts are more qualified than their minority counterparts. However, the proportion of qualified jurors from anglo dominant census tracts was the same as that of the judicial district as a whole.

We suggested an alternative sampling strategy of random selection of census tracts which provides a representative list of tracts from which potential jurors could be summoned. Our simulation analysis showed that selected census tracts could provide a list of potential jurors that would be unbiased, i.e., racially representative. That is, the impanelment lists would have a racial composition similar to the judicial district. Our method of randomly selecting census tracts is clearly superior to the selection method currently employed in Los Angeles County, because the potential jurors coming from the selected tracts are evenly distributed and have an equal chance of being selected. The random selection of census tracts, thus, is congruent with requirements established by both the Federal Jury Selection and Service Act in 1968 and the California Code of Civil Procedure in 1981.[9]

[9] Federal Jury Selection and Service Act was passed in 1968 to guarantee that "all litigants in Federal courts entitled to trial by jury shall have the right to grand and petit juries selected

Bibliography

Alker, Hayward R., Jr. & Joseph J. Barnard. 1978. Procedural and social biases in the jury selection process. The Justice system journal 3:220–241.

Alker, Hayward R., Jr., Carl Hosticka Mitchell. 1976. Jury selection as a biased social process. The law and society review 9:9–41.

Butler, Edgar W. 1980a. Torrance Superior Court panels and population analysis. University of California, Riverside.

Butler, Edgar W. 1980b. Van Nuys Superior Court panels and population analysis: May 7, 1979 yhrough September 24, 1979. University of California, Riverside.

Butler, Edgar W. 1981. The 1980 Los Angeles County jury selection study: Compton Superior Court. University of California, Riverside.

Butler, Edgar W. and Hiroshi Fukurai. 1984. An evaluation of jury panel selection procedures: the North Valley Superior Court, Los Angeles County. University of California, Riverside.

Cherigny, Paul G. 1975. The Attica case: a successful jury challenge in Northern City. Criminal law bulletin 11:157–172.

Cullen, Francis T. & Bruce G. Link. 1980. Crime as an occupation. Criminology 18:399–410.

de Cani, John S. 1974. Statistical evidence in jury discrimination cases. The journal of criminal law and criminology 65: 234–238.

Fukurai, Hiroshi. 1985. Institutionalized racial inequality: a theoretical and empirical examination of the jury selection process. Unpublished dissertation, University of California, Riverside.

Fukurai, Hiroshi & Edgar W. Butler. 1986. Assimilation and internal colonialism models of jury selection. [forthcoming]

Fukurai, Hiroshi & Edgar W. Butler. 1986. The jury selection process: institutionalized inequality. [forthcoming]

Hepburn, John R. 1978. Race and the decision to arrest: an analysis of warrants issued. Journal of research in crime and delinquency 15:54–73.

Heyns, Barbara. 1979. 1979 jury analysis. (Superior Court, County of Los Angeles, No. A–344097, Joseph Piazza, defendant.)

Huebner–Dimitrius, Jo–Ellan. 1984. The representative jury: fact or fallacy? Unpublished Dissertation. Claremont Graduate School.

Kairys, David. 1972. Juror selection: the law, a mathematical method of analysis, and a case study. American criminal law review 12:771–806.

Kairys, David, Joseph B. Kadane, & John P. Lehoczky. 1977. Jury representativeness: a mandate for multiple source lists. California law review 65:776–827.

[9](cont'd) at random from a fair cross section of the community in the district or division wherein the court convenes" (U.S. 1968, Section 1861).

Kramer, Ronald C. 1982. From 'habitual offenders' to career criminals'. Law and human behavior 6:273-293.

Levine, Adeline Gordon & Claudine Schweber-Koren. 1976. Jury selection in Erie County: changing a sexist system. Law and society review11:43-55.

Moore, Lloyd E. 1973. The jury: tool of kings, palladium of liberty. Cincinnati: The W.H. Anderson Company.

Turk, Austin T. 1981. The meaning of criminality in South Africa. International journal of sociology and law 9:123-135.

U.S. 90th Congress Senate Report. 1967. No. 891.

U.S. 90th Congress House Report. 1968. No. 1076.

Van Dyke, Jon M. 1977. Jury selection procedure. Massachusetts: Ballinger Publishing Company.

anon. 1970. The case for black juries. Yale law journal 79:531-550.

Cases Cited

Alexander v. Louisiana, 405 U.S. 625 (1972)

City of Mobile, Ala v. Bolden, 466 U.S. 55 (1980)

Duren v. Missouri, 439 U.S. 357 (1979)

People v. Estrada, 155 Cal. Rptr. 731 (1979)

People v. Graham, 160 Cal. Rptr. 10 (1979)

People v. Harris, 36 Cal. 3d 36, 201 Cal. Fprt. 782 679 P. 2d 433 (1984)

People v. King, 49 Cal. Rptr. 562 (1966)

People v. Sirhan, 7 Cal. 3d 258 (1978)

People v. Wheeler, 148 Cal. Rptr. 890 (1978)

People v. White, 43 Cal. 3d 740 (1954)

Peters v. Kiff, 407 U.S. 493 (1972)

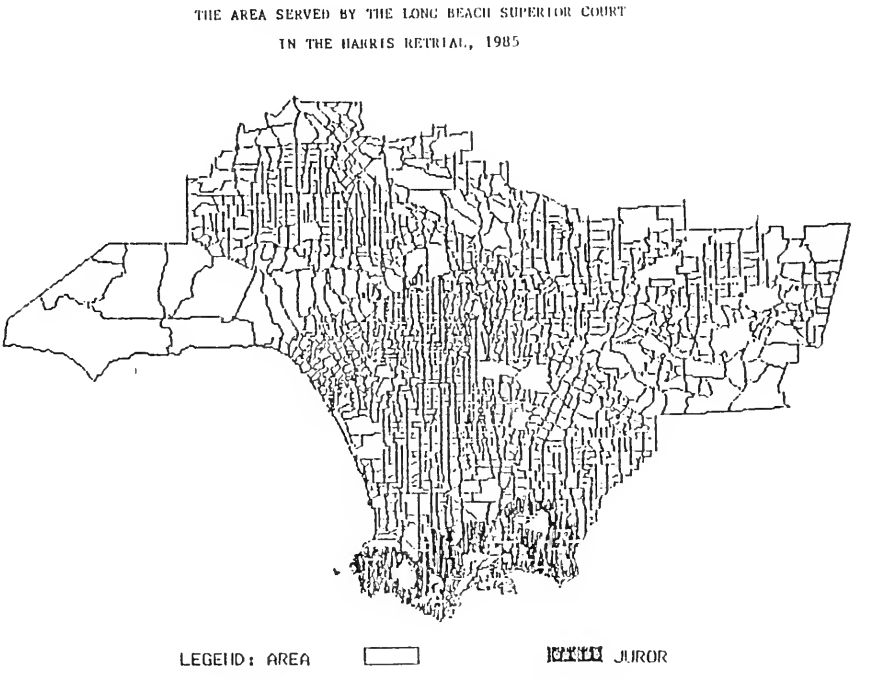
State v. Holstrom, 43 Wis. 465, 168 N.W. 2d 574 (1969)

Taylor v. Louisiana, 419 U.S. 522 (1975)

Thiel v. Southern Pacific Co., 328 U.S. 217 (1946)

United States v. Fernandez, 480 F. 2d 726, 732-33 (2d Cir. 1973).

Map A


THE AREA SERVED BY THE LONG BEACH SUPERIOR COURT
IN THE HARRIS RETRIAL, 1985
LEGEND: AREA
JUROR

MAP A

Figure 1

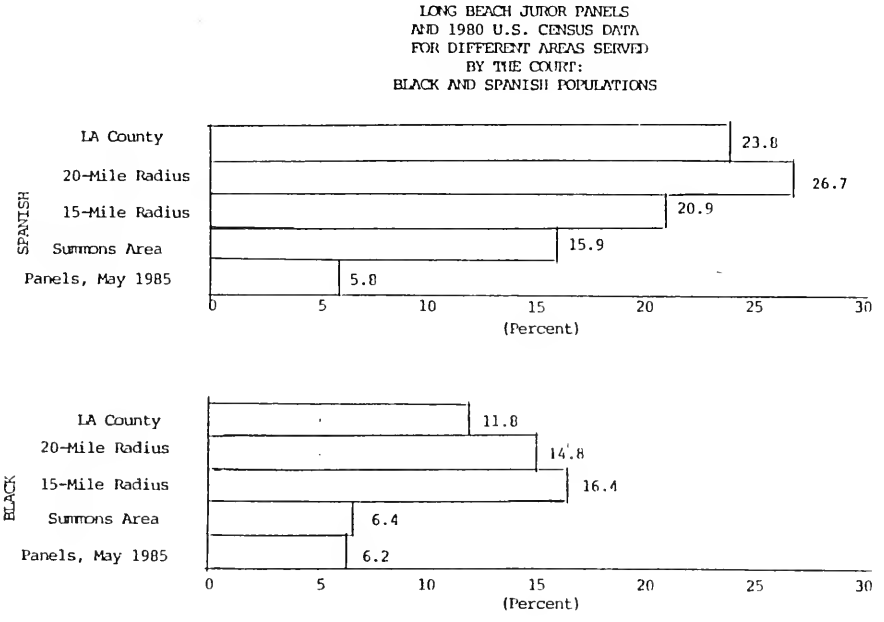

LONG BEACH JUROR PANELS
AND 1980 U.S. CENSUS DATA
FOR DIFFERENT AREAS SERVED
BY THE COURT:
BLACK AND SPANISH POPULATIONS
SPANISH
LA County
20-Mile Radius
15-Mile Radius
Summons Area
Panels, May 1985
23.8
26.7
20.9
15.9
5.8
0
5
10
15
20
25
30
(Percent)
BLACK
LA County
20-Mile Radius
15-Mile Radius
Summons Area
Panels, May 1985
11.8
14.8
16.4
6.4
6.2
0
5
10
15
20
25
30
(Percent)

Table 1

TABLE 1

LONG BEACH SUPERIOR COURT ELIGIBLE POPULATION AND IMPANELMENT LISTS

Variable	Long Beach District 20-Mile Radius	Impanelment* Lists	Absolute Disparity	Relative Disparity	Z Score	Chi-Square Value
Non-Hispanic	79.2%	89.9%	10.8	13.7	7.5**	11.7
Hispanic	20.9%	10.1%	-10.8	-51.7	-7.5**	11.3

* 4/24/85 through 6/12/85

** Significant at <<<0.0001 level.

Table 2

TABLE 1

LONG BEACH: 8 PANELS
APRIL 24, 1985 TO JUNE 12, 1985

Panel	Date	Hispanic			Black	
		Impanelment List	Census Tract	Z Score	Census Tract	Z Score
1.	4-24-85	9.0%	10.7%	-2.9	7.0%	-2.5
2.	5-1-85	10.0	16.1	-2.7	7.7	-2.3
3.	5-8-85	13.0	12.5	-1.9	6.3	-2.7
4.	5-15-85	11.0	11.8	-2.4	7.8	-2.3
5.	5-22-85	12.0	14.2	-2.2	5.2	-3.0
6.	5-29-85	11.0	13.7	-2.4	4.7	-3.2
7.	6-5-85	7.2	12.2	-3.8	9.2	-2.2
8.	6-12-85	8.0	12.7	-2.7	5.2	-3.6
	TOTAL	10.1	15.9 *	-7.5	6.4 *	-7.8

* Percentages are calculated on the basis of all included census tracts.

Table 3

TABLE 3

REPRESENTATION OF CENSUS TRACTS ON 8 PANELS
LONG BEACH SUPERIOR COURT DISTRICT
APRIL 24, 1985 THROUGH JUNE 12, 1985

NO. OF TIMES ON 8 PANELS	NO. OF CENSUS TRACTS FREQUENCY	PERCENT
1	10	9.34
2	8	7.47
3	11	10.28
4	9	8.41
5	13	12.15
6	8	7.47
7	9	8.41
8	8	7.47
9	2	1.86
10	3	2.80
11	3	2.80
12	8	5.60
13	7	6.54
14	1	0.93
17	1	0.93
18	2	1.86
19	1	0.93
20	1	0.93
22	1	0.93
23	1	0.93
34	1	0.93
36	1	0.93

Median = 6

Mean = 7.5

Table 4

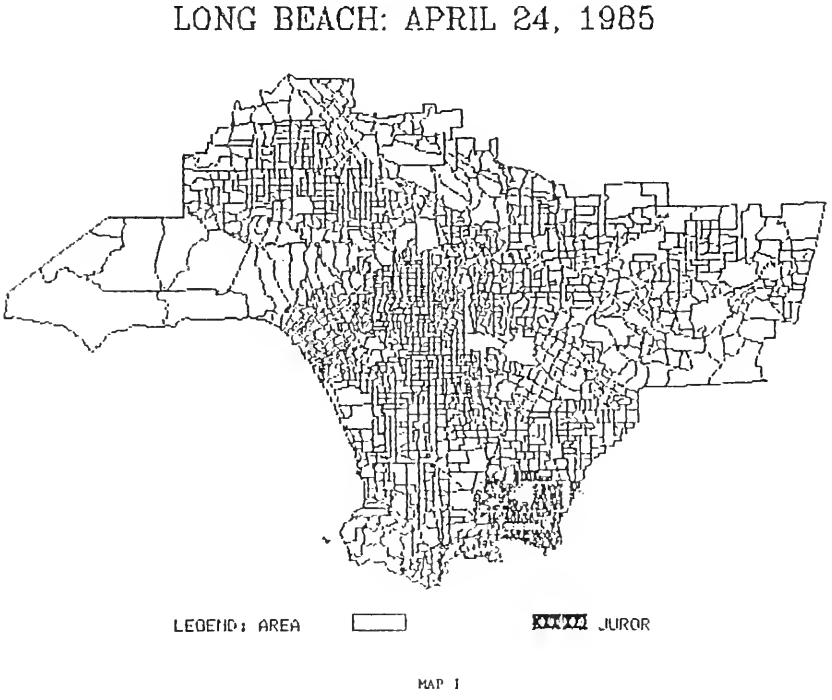
TABLE -

AVERAGE REPRESENTATION OF CENSUS TRACTS ON 8 PANELS
LONG BEACH SUPERIOR COURT DISTRICT

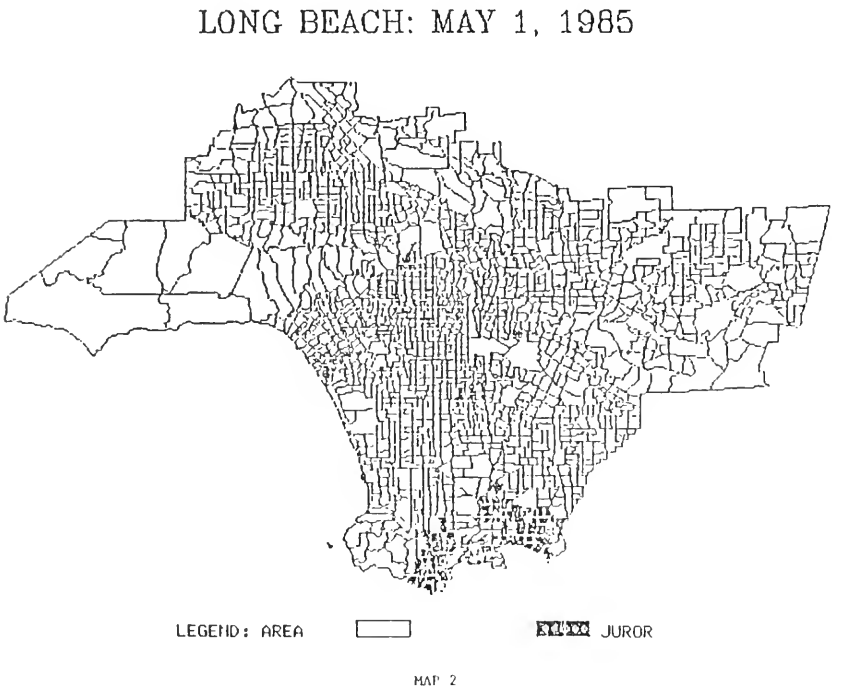
Race	Long Beach* District	Mean		Median	
		8 Times or More	7 Times or Less	7 Times or More	6 Times or Less
Black	16.4%	4.3%	7.6%	4.6%	8.3%
Hispanic	20.9	9.2	19.5	9.2	22.9

* 430 census tracts included

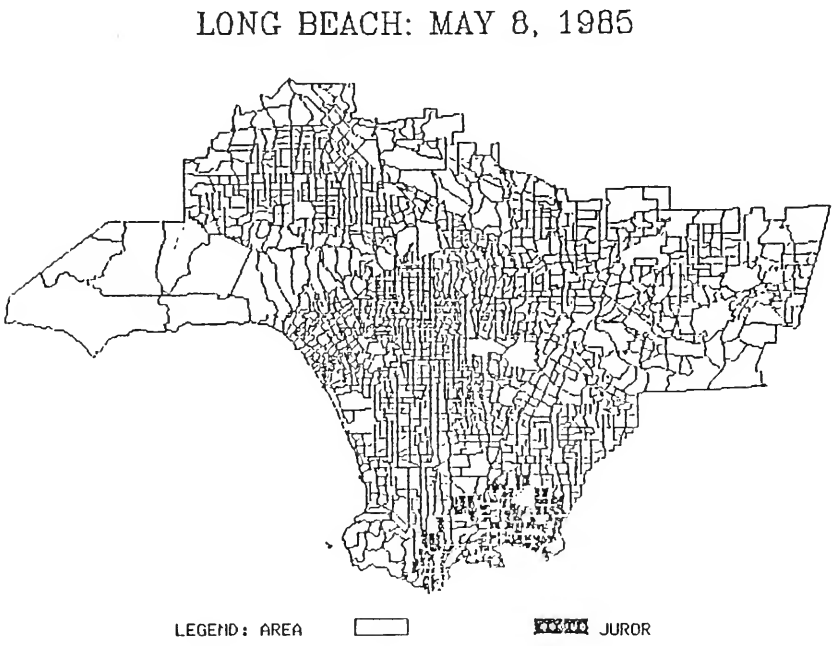
Map 1



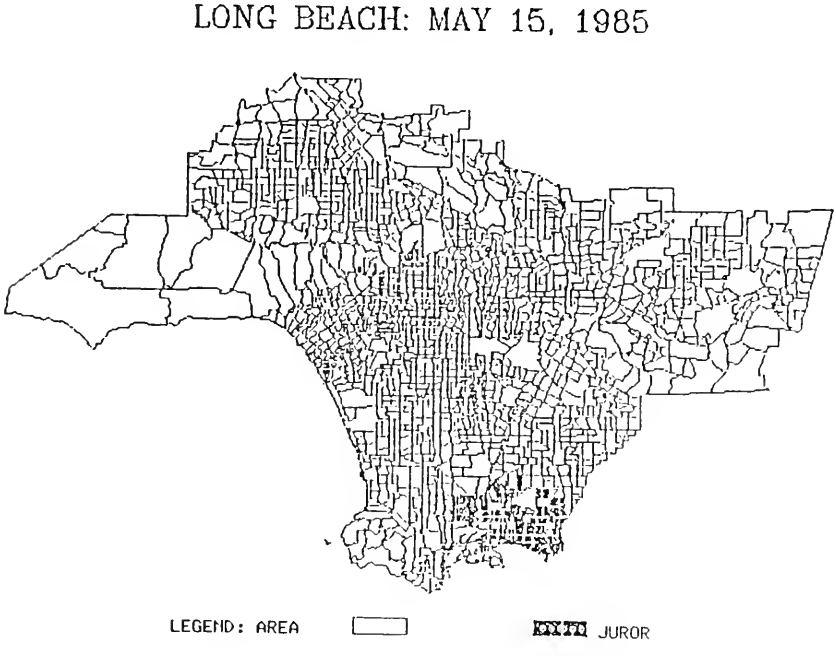
Map 2



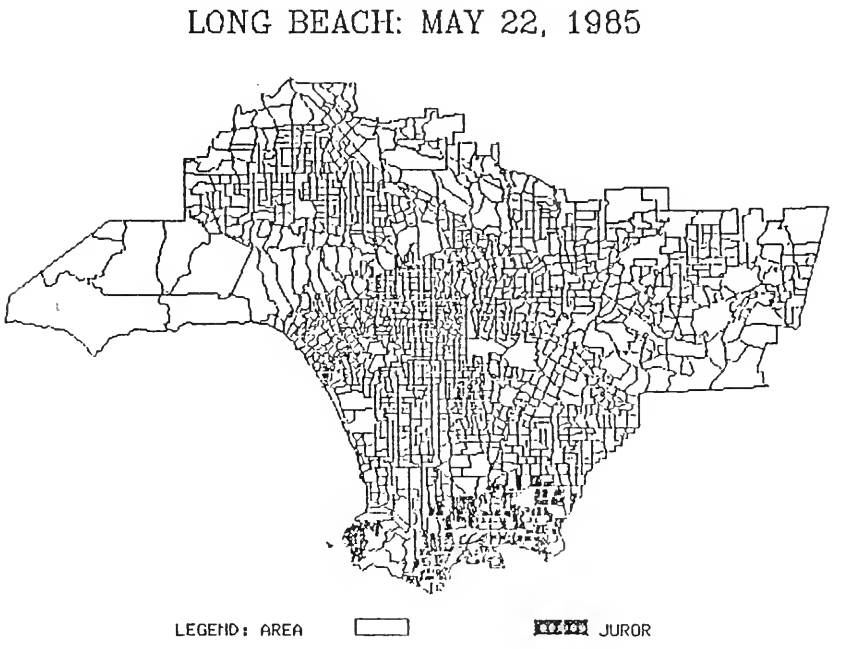
Map 3



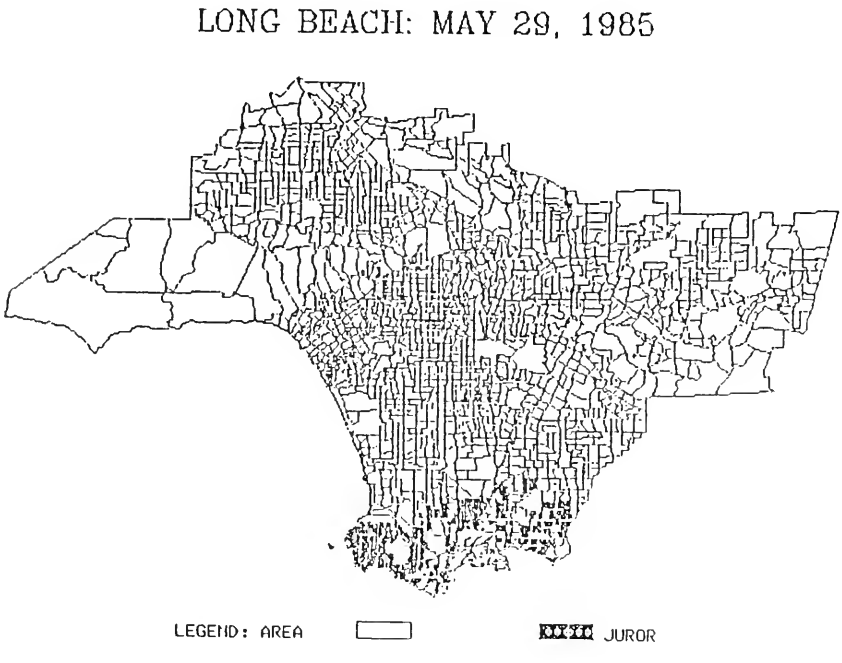
Map 4



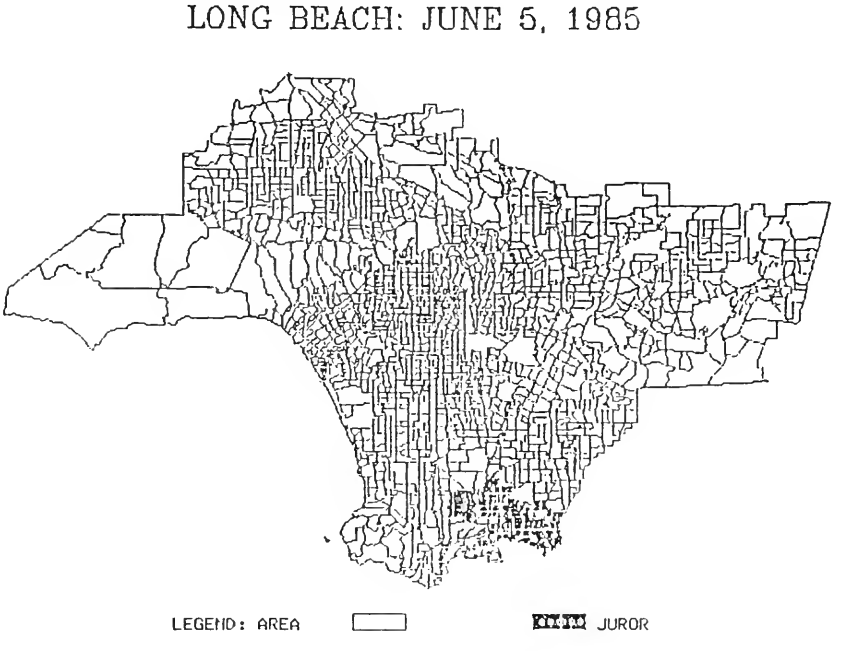
Map 5



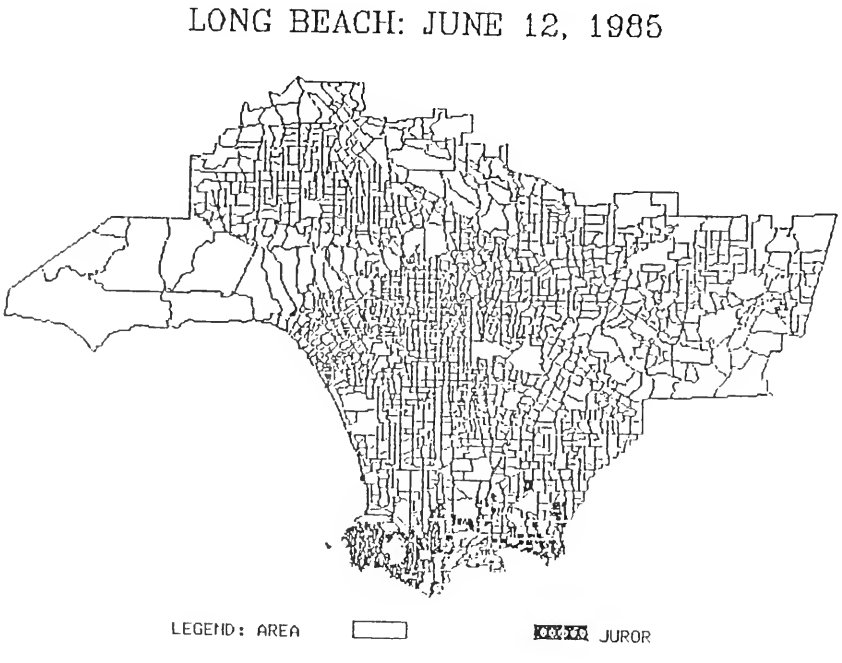
Map 6



Map 7



Map 8



MAP 8

Table 5

TABLE 5

QUALIFIED JURORS*
LONG BEACH SUPERIOR COURT

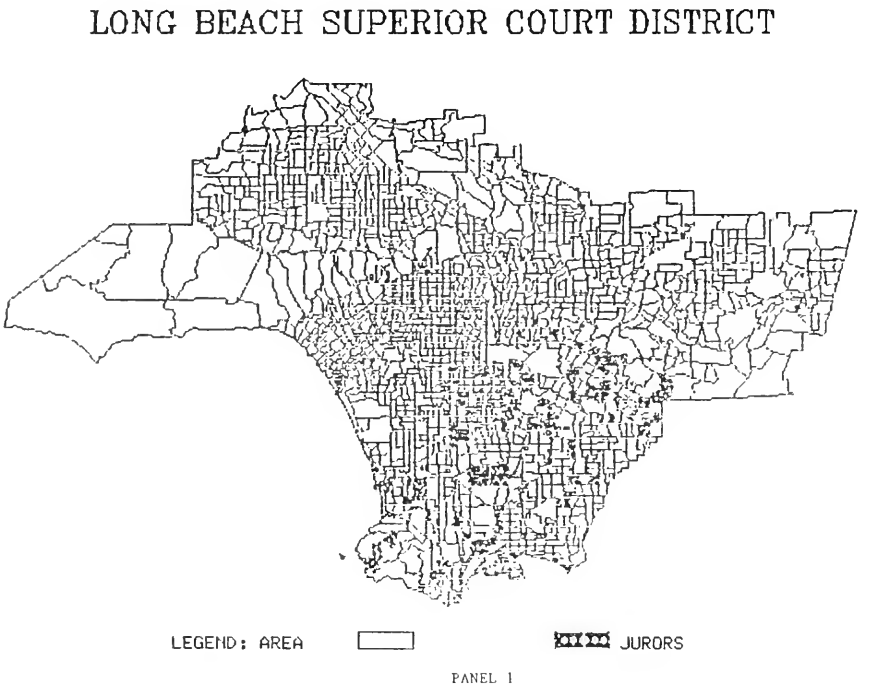
	Long Beach Judicial District[1]		Impanelment Lists[2]	
	No.	Percent	No.	Percent
Total Jurors	243,274	100%	62,758	100%
Qualified Jurors	46,436	19.1%	12,459	19.9%

1. Total number of census tracts are 439.

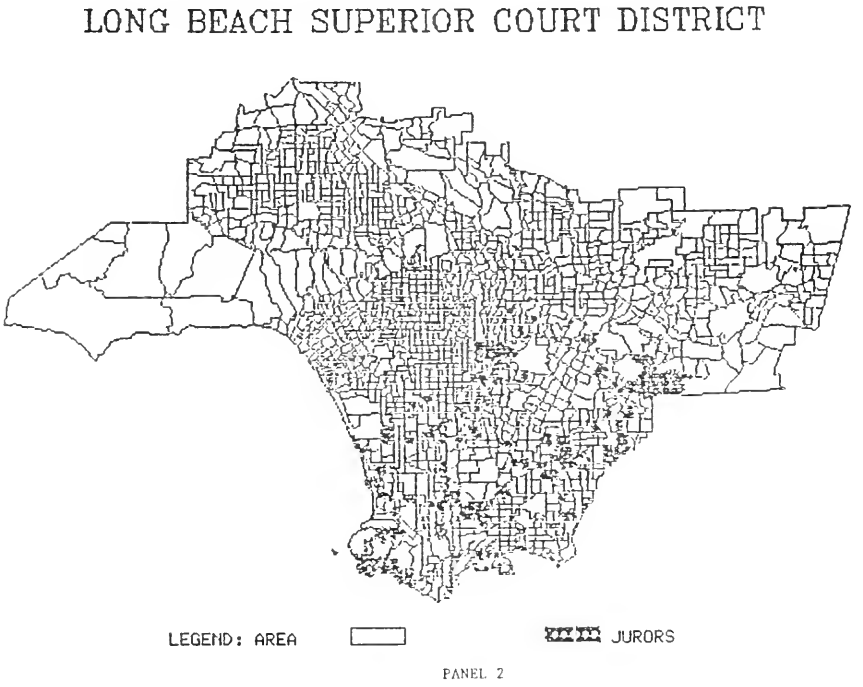
2. Total number of census tracts are 107.

* Source: Los Angeles Jury Supervisor Ray Arce and his computer consultants, 1985

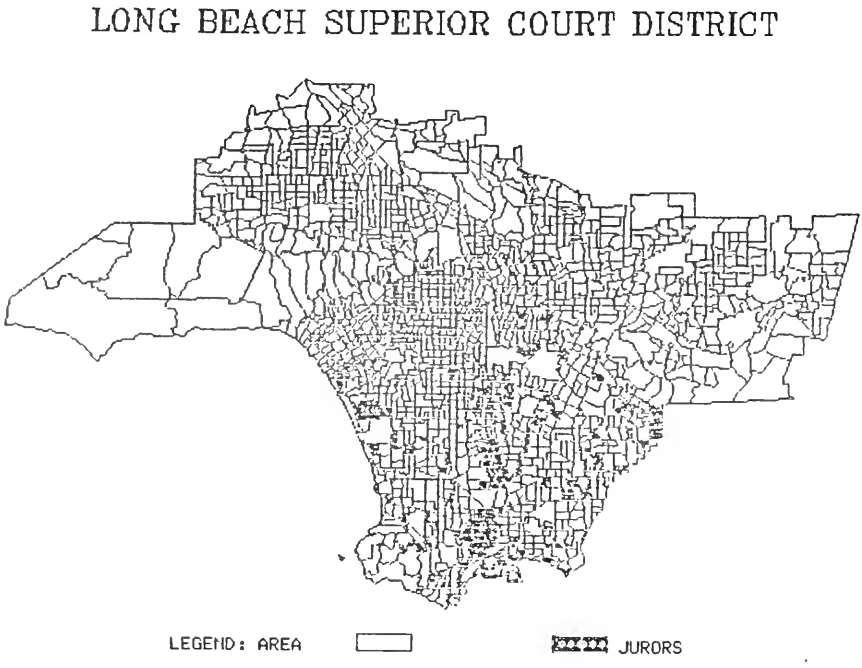
Map 9


LONG BEACH SUPERIOR COURT DISTRICT
LEGEND: AREA
JURORS
PANEL 1
MAP 9

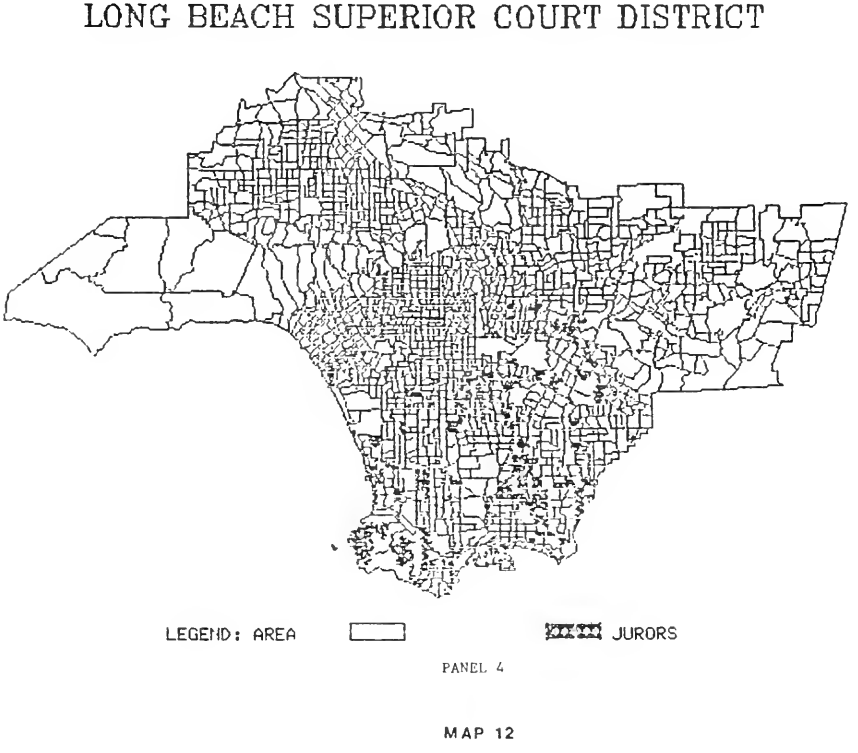
Map 10


LONG BEACH SUPERIOR COURT DISTRICT
LEGEND: AREA
JURORS
PANEL 2
MAP 10

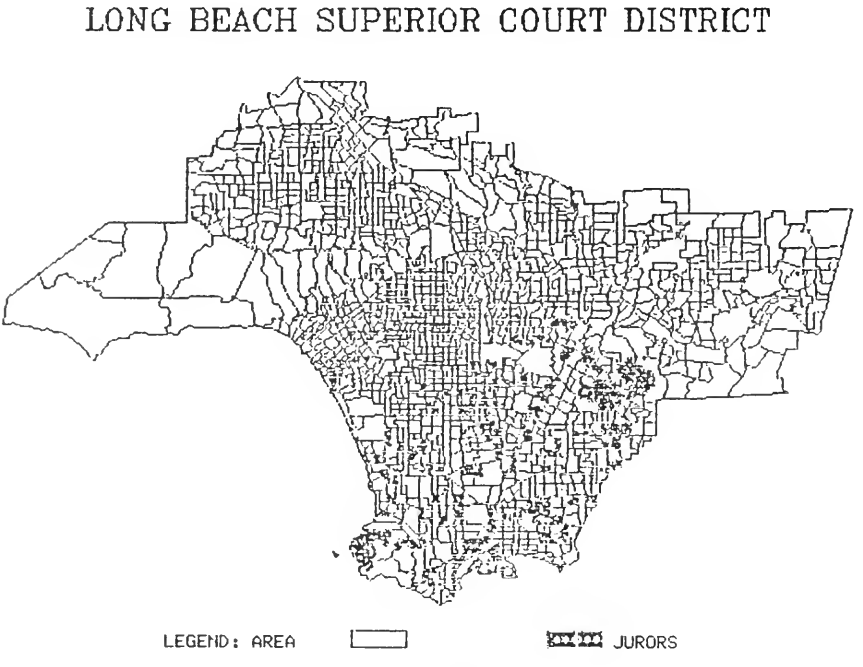
Map 11


LONG BEACH SUPERIOR COURT DISTRICT
LEGEND: AREA
JURORS
PANEL 3
MAP 11

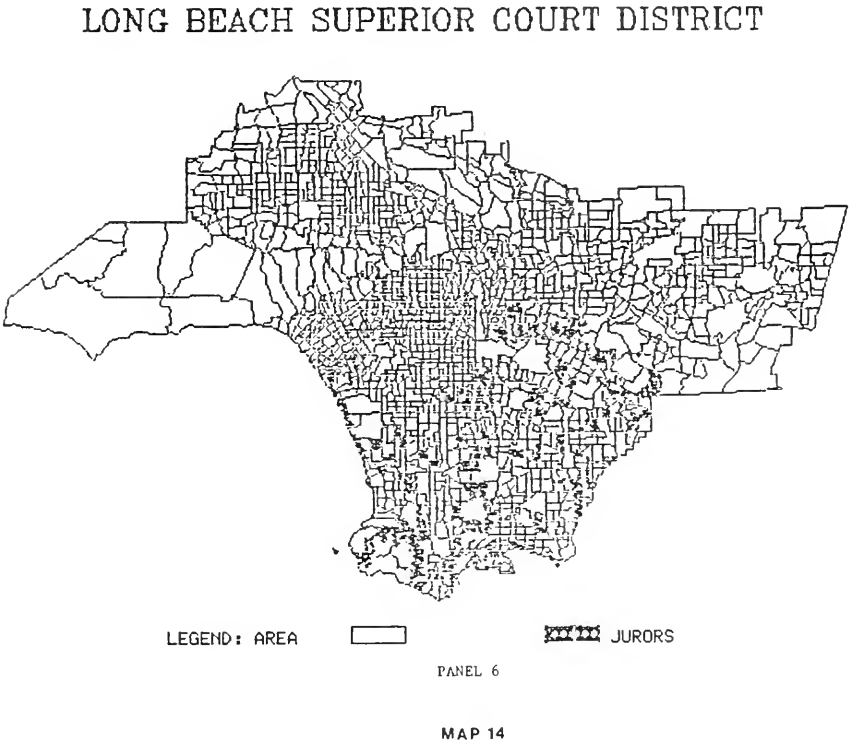
Map 12


LONG BEACH SUPERIOR COURT DISTRICT
LEGEND: AREA
JURORS
PANEL 4
MAP 12

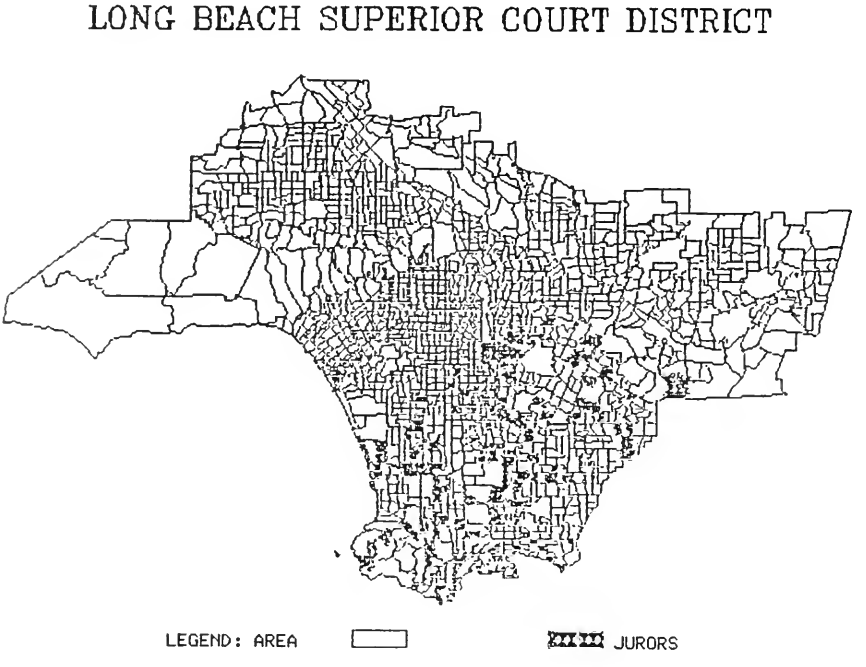
Map 13


LONG BEACH SUPERIOR COURT DISTRICT
LEGEND: AREA
JURORS
PANEL 5
MAP 13

Map 14


LONG BEACH SUPERIOR COURT DISTRICT
LEGEND: AREA
JURORS
PANEL 6
MAP 14

Map 15


LONG BEACH SUPERIOR COURT DISTRICT
LEGEND: AREA
JURORS
PANEL 7
MAP 15

Map 16

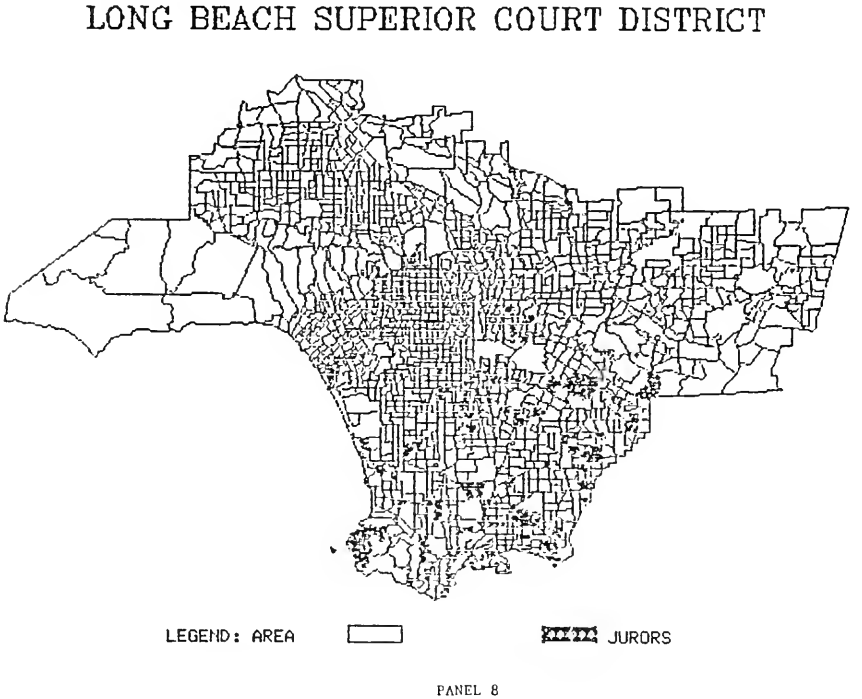

LONG BEACH SUPERIOR COURT DISTRICT
LEGEND: AREA
JURORS
PANEL 8
MAP 16

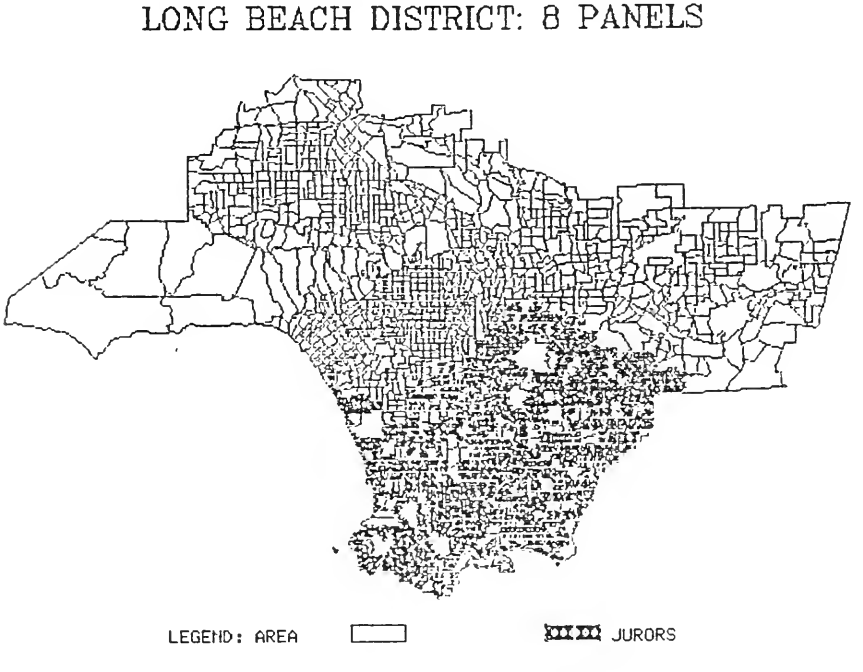
Table 6

TABLE 6

LONG BEACH: 8 PANELS BY RANDOM SELECTION
APRIL 24, 1985 TO JUNE 12, 1985

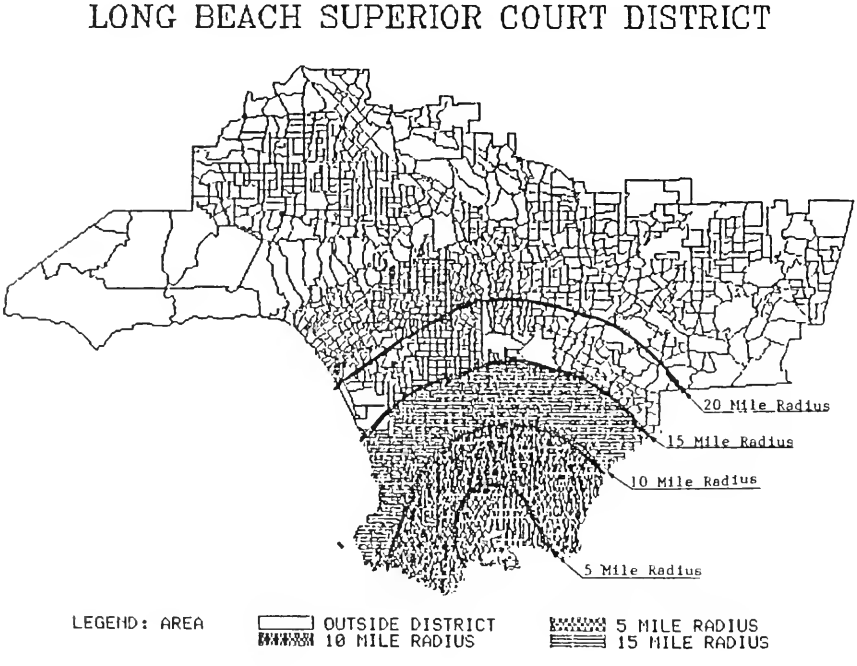
Panel Date	Hispanic		Black	
	Census Tract	Z Scores	Census Tract	Z Scores
1. 4-24-85	26.7%	0.0	17.1%	0.6
2. 5-1-85	23.9	-0.6	12.6	-0.6
3. 5-8-85	29.8	0.7	15.9	0.3
4. 5-15-85	24.6	-0.5	13.9	-0.3
5. 5-22-85	24.7	-0.5	11.3	-1.0
6. 5-29-85	30.2	0.8	8.7	-1.7
7. 6-5-85	25.8	-0.2	13.8	-0.4
8. 6-12-85	25.9	-0.2	12.0	-0.7
TOTAL	27.4	0.4	13.7	-0.9

Map 17



MAP 17

Appendix A


LONG BEACH SUPERIOR COURT DISTRICT
20 Mile Radius
15 Mile Radius
10 Mile Radius
5 Mile Radius
LEGEND: AREA
OUTSIDE DISTRICT
10 MILE RADIUS
5 MILE RADIUS
15 MILE RADIUS
APPENDIX A

CD-ROM and the data archive: beyond retrieval

by Ann Gerken[1]
University of California, Berkeley

Optical storage offers data archives a new medium for storing data and a change in the information retrieval environment. As with any new technological development, it is important to evaluate the impact and tradeoffs of adapting to new equipment and systems. By stepping back for an overview of information storage and access, better evaluations may be made within the context of the "big picture" of data use. Before embracing this new technology, it is important to understand where we are going and how optical storage can help us. Optical media offer some solutions to storage problems, but have not yet shown themselves to be solutions to the overall needs of information providers and users.

First of all, thanks go to Forrest Williams, Census Bureau, and to George Hall and Courtney Slater, of Slater-Hall Information Products, and to Ed Spar, of Market Statistics. These people have provided us with CDROM test disks for our evaluation, and I sincerely appreciate their support. Some of the points I will discuss today have been included in the report of the State Data Center's Subcommittee on CDROM. This report is available from John Kavalaunis of the Data User Services Division. My ideas do not necessarily reflect all the opinions of that subcommittee.

Consider the advantages of optical media as storage devices. CDROM and WORM disks provide a minimum of 10 years' secure storage. The disks provide random access, and are portable and compact. However, there are some problems. For most data archives who rely upon large centralized computing systems, the shift from storing data on tapes and disk packs to storing data on microcomputer peripherals means a dramatic shift in responsibility. How will data archives provide multiple user access to data on microcomputers? Can archives handle the equipment and network burden? Another problem with CDROM is that archives cannot produce CDROM disk copies in house; there is no local recording capability. This

[1]Presented to The Association of Public Data Users October 22, 1987, Washington, DC

means that if data archives are to produce optical backups of their holdings, they must use WORM disks, and will have to acquire and maintain two kinds of optical devices. Other problems are access speed, and in some cases, limited storage capability (CDROMS only hold 4 reels of tape, and STF4 for California takes up 14 reels of tape). Another issue to consider is that the software to access the data on CDROMs and WORMs has not kept pace with the technology. Even though information on microcomputers appears to be closer to the end user, the interfaces and support for access are often limited. And finally, we must ask when will optical storage be cost effective, and when will the major data distributors begin to use this technology?

Information retrieval is the process of extracting specific data items from specific records in a data file. At this time, information retrieval software is produced for each CDROM product since standard retrieval software for statistical data files is not yet available. At its best, this retrieval process is informed, flexible, logical, and fast. However, since each data file comes with its own custom software there is a wide range in the quality of these software interfaces. Users will be required to learn multiple retrieval processes. Also, statistical CDROM products are more expensive to produce since producing custom software for each data file requires extensive programming.

The premastering and production of the optical disks themselves are different from producing magnetic tape. The physical process itself is more expensive, and the nature of the files on the disk differ. Since the medium is randomly accessed, "disk geography," or where on the disk various records will be placed, is more important than with tapes. Disk geography can improve performance if done carefully. Disk indexes should be produced and mastered with the data file if the disk is to prove efficient in sophisticated retrieval applications. Error detection and correction codes must also be provided. In summary, "raw" data files on CDROM should be more carefully developed than those provided on magnetic tape. To exploit the technology and access software, optical products should be distributed with value added files.

After information is retrieved, "beyond retrieval", lies the world of post processing where the information is manipulated and packaged to user specifications. At its best, this processing should be supported by an integrated system providing clear steps for users to follow. Information retrieved from CDROM products should be compatible with inhouse software, and transfer of the information should be standard and simple.

In our evaluations, we considered three CDROM products in relation to data product preparation, information retrieval, and post processing. The three CDROM products we evaluated were: Census Test Disk #1, which includes data and software for ZIP codes, population estimates, and the 1982 Census of Agriculture; Slater Hall's 1982 Census of Agriculture disk; and "Your Marketing Consultant" form Market Statistics. The Census disk offers simple extract programs that demonstrate the usefulness of optical storage. The procedures for retrieval are straight forward and can be done by any novice computer user. A highlight of the Slater Hall software is its online help for each variable in the data base, as well as its numeric search capabilities. "Your Market Consultant" offers the capability of sorting data by user defined criteria, and allows users to add their own data to the tables.

Our evaluation at the State Data Program emphasized the retrieval and post processing capabilities of these products, but did not take a close look at the premastering and production of the CDROMs. The issues of disk geography and indexing are yet to be evaluated. However, in the area of retrieval, it is clear that a standard procedure has emerged among these

products. Users progress through a series of simple procedures, first selecting geographic areas, then identifying a table or variables, then extracting data from the disk, and finally producing display, report or file output.

What more do archives and users need from this kind of product? Given that a CDROM product provides the information that a user wants, a separate procedure is often required after retrieving information. A user must sort and create custom tables, graphics, or spread sheets in another system. Obviously, there is a need to coordinate software compatibility and command language among in-house software, so that users are not faced with a confusing array of software and transfer methods. In addition, the lack of numeric manipulation and statistical analysis in the CDROM software requires an interface with statistical software currently in use. Data as well as textual information must be moved in efficient ways and in compatible formats.

None of the current CDROM products contain microdata, i.e. data for individual households, persons or establishments. The data that are available are summaries for geographic units, useful for many applications, but are only part of the larger universe of data. Extracts of microdata files are regularly produced in the archive environment, and products and procedures using CDROM technology should be considered.

Our suggestions are that 1) data distributors keep an open mind about which kind of optical media to adopt, so that archives are not forced to develop dual systems, one for data acquisition (CDROM) and one for data file backup (WORM). We also are concerned about the technology becoming outdated. 2) Optical products should be mastered to allow for optimal retrieval by a variety of software, including relational database management software. 3) Retrieval software should provide output that is compatible with other software, and the transfer of data should be an integral part of the software product. 4) Data distributors should view their products not only as stand-alone retrieval systems, but as pieces of a larger information support system.

Some questions for the future: How can optical storage be integrated into the existing information fabric? Who will build the interfaces for informed transfer of data from system to system? Do we place responsibility upon the private sector to develop retrieval products, and how involved can we become? What information systems are in use that identify, retrieve, transfer, and analyze data? How may these systems influence the development of CDROM and WORM products?

Data archives have a long history of responding to technological changes: from IBM cards to low density tapes, from floppy diskettes and random access disk packs to optical media. Eric Tanenbaum wrote: "Archives have a privileged role among information providers for they were among the first to computerize. Thus they offer a rare perspective from which to view the changes." (IASSIST Quarterly, Spring 1986) I believe that archivists will be studying these changes and will offer a unique perspective in emphasizing the integration of new technology into existing services and procedures. Archivists and other information professionals have an opportunity to participate in the development of standard, reliable storage media, to assist in the design and evaluation of powerful flexible retrieval software, and to help create dynamic post-retrieval environments.□

Contents of current journals

European political data newsletter
nr. 64 September 1987

Bits & bytes review
vol. 1(5), April 1987

Bits & bytes review
vol. 1(6), May 1987

ESRC data archive bulletin
nr. 38 September 1987

Scope: humanities computing update
vol. 5(4), Jul-Aug 1987

Survey methodology
vol. 13(1), June 1987

Data user news from the Bureau of the Census
vol 22(9), Sept., 1987

Data user news from the Bureau of the Census
vol 22(11), Nov., 1987

Communications of the A.C.M.
V.30(11) November 1987

Articles:

Social science microcomputer review
V.5(4) Winter 1987

Machine Readable Records bulletin
vol. 5(3), Fall 1987

Computational statistics & data analysis
V.6(1) January 1988

Scope: humanities computing update
vol. 5(5), Sept.-Oct., 1987

Communications of the A.C.M.
V.31(1) January 1988

Data user news from the Bureau of the Census
vol. 22(12), Dec. 1987

Electoral studies
V.6(3) December 1987

ICPSR Bulletin
(Inter-University Consortium for Political and Social Research)
Dec. 1987

ESRC data archive bulletin
N.39 January 1988

News:

New acquisitions:

Research organisations, data institutions and foreign archives:

Software bulletin:
Notes:

Books:

Book notes:

Appendices:

Communications of the ACM
V.31 N.2 February 1988

Articles:

Computing practices:

Research contributions:

Departments:

European political data newsletter
No.65 December 1987

NSD brukermelding
(Norwegian Social Science Data Service)
1988:1, February 1988

NSD brukermelding
(Norwegian Social Science Data Service)
1988:2, March 1988

Bits & bytes review
vol. 1(7), June 1987

SSD kontakt
(Swedish Social Science Data Service)
1-2/1987

DDA nyt
(Danish Data Archives) nr. 42, summer 1987

DDA-0658 Danish election studies, continuity file 1971-1981.
DDA-1052 OBSERVA political index polls, 1984.
DDA-1053-1064 OBSERVA political index polls, Jan.-Dec. 1984.
DDA-1069 Experienced workers and new technology, 1983.

p. 49 New data files: [in Danish]
DDA-1237 Danish leisure study, 1987: background variables.
DDA-1242 Danish leisure study, 1987: adult questionnaire.
DDA-1243 Danish leisure study, 1987: adult telephone interviews.
DDA-1244 Danish leisure study, 1987: daily time use.
DDA-1245 Danish leisure study, 1987: child questionnaire.
DDA-1246 Danish leisure study, 1987: child interviews.
DDA-1347 Danish omnibus survey, January 1987.
DDA-1348 Danish omnibus survey, April 1987.
DDA-1349 Social networks and care of old people, 1987-1988.
DDA-1350 Maltreatment of old people - failing care, 1987-1988
. DDA-1351 Development of trade unions in Europe after World War II.
DDA-1347 Danish omnibus survey, January 1987.

p. 57 English summary.
p. 73 Order form. [in Danish]

Humanistiske data
(Norwegian Computing Centre for the Humanities)
3-87

ZA Information
(Zentralarchiv fuer Empirische Sozialforschung)
nr. 21, November 1987

News and notes:

ZUMA Nachrichten
(Zentrum fuer Umfragen, Methoden und Analysen e.V.)
nr. 21, November 1987

Historical social research
nr. 42/43, April/July 1987

Special issue: the multivariate analysis of nominal and ordinal data.

Historical social research
nr. 44, October 1987

Special issue: history of employment structure in 20th century Europe.

Current research

NOTICE

If your archive has a written collection development policy for machine-readable data files, please send a copy to, or contact Dan Tsang, Main Library, University of California, PO Box 19557, Irvine CA 92713, USA. Tel: (714) 856-4978 Email: dtsang@orion.cf.uci.edu

He is researching such policies at academic libraries.

Australian election survey, 1987

Background

This survey of 1,825 respondents is the first purpose–designed national election study undertaken in Australia. It has two goals. The first is to continue the broad line of inquiry established by the 1967 and 1979 Australian National Political Attitudes Surveys into the long–term character of national political attitudes and behaviour. The second, and more novel, is to focus more clearly on short–term forces, like issues and leaders, specific to the 1987 election. The principal investigators are Roger Jones of the Australian National University, Ian McAllister of the University of New South Wales and Anthony Mughan formerly of the Australian National University and now of the Ohio State University. The survey was funded by the Australian National University and the University of New South Wales.

The Survey

A systematic random sample of 3,061 respondents was drawn from the electoral rolls at the Australian Electoral Commission two weeks before the election of 11 July 1987. The self–completion (mail out, mail back) questionnaire comprises 195 variables divided into six broad sections: The Federal Election, Political Leaders, Election Issues, Social and Political Goals, Education and Work and Personal Background.

The Federal Election section contains questions on concern about the election outcome, the influence of the media, party identification, vote in the 1984 and 1987 elections, the timing of the voting decision and the importance to the vote of a number of issues.

The Political Leaders section enquiries about feelings towards the party leaders indicated by ratings from 0 to 10, about their leadership qualities and their effectiveness as Prime Minister. It also assesses perceptions of the unity, moderation and class appeal of the Labour, Liberal and National parties.

The Election Issues section is the longest, focussing on economic voting, privatisation, government spending, trade union and big business power, uranium mining, law and order, Asian migration, marijuana, abortion, aborigines, testing for AIDS, censorship, homosexuality, the environment and

opportunities for women.

The Social and Political Goals section contains Inglehart's postmaterialism questions, trust in government and modes of political participation questions.

Education looks at number of years of schooling, state or Catholic school and qualifications obtained. Work asks about employment status, occupation, public or private employment, supervisory responsibilities, trade union membership and subjective class.

The personal background questions include sex, place and length of residence, age, place of birth, date of arrival in Australia, father's occupation and party preference, marital status and children, spouse's occupation, trade union membership and party pareference and the respondent's religion and religious practice.

Availability

The data file and codebook are now available from The Social Science Data Archives, The Australian National University, GPO Box 4, Canberra, ACT 2601, Australia. Telephone (062) 494400.

Career opportunities

Job Opening

Position Title: **Associate Director, Government Data Center**
Harvard University

The Government Data Center (GDC) is the home of the Government Department's microcomputer lab and Harvard's primary computerized social science data and codebook archive. Since the director of the GDC is a member of the Government Department faculty, the associate director is the top staff position at the GDC and is highly visible within the Harvard community. The occupant of this position provides technical advice and guidance to faculty, staff, and students on computational problems related to coursework, research, and administration. Serves as consultant to other related social science departments, and acts as the department's liaison with other departments and units of the University on computer-related issues. Supervises the acquisition and cataloging of data sets and codebooks, and assists in developing and implementing the GDC's financial and other policies. With his or her staff, the associate director has considerable flexibility. Depending on skills and interests, the associate director is welcome to choose special projects to work on; these might include: writing computer-aided instruction software, automating GDC operations, teaching mini-courses on computer software and hardware, assisting instructors in classroom teaching related to computers, or writing documentation. Other ideas and special projects may be developed by the new associate director.

Bachelor's degree required; advanced degree preferred. Applicants should have working knowledge of IBM microcomputers and associated software. Experience with word processing and statistical software is preferred. Familiarity with one or more programming languages (BASIC, Fortran, Pascal, etc.) and mainframe operating systems are helpful. Verbal, administrative, and interpersonal skills are essential.

Salary may be in low to high $30's, depending on qualifications. Interviews begin immediately. Please send a vitae with references as soon as possible to:

Professor Gary King
Department of Government
Harvard University
Littauer Center M37
Cambridge, MA 02138

Phone: 617-495-2027
BITNET: gmkHarvunxw
Internet: gmkWjh12.harvard.edu
UUCPnet: gmkWjh12

Course offering:
Management of computer-readable files

The University of British Columbia School of Library, Archival and Information Studies is offering the following Extra Sessional course in the Summer Session. Registration is from March 12 - June 28.

Course: LIBR 651B - Management of computer-readable files

Instructor: Laine G.M. Ruus

Objectives: to acquire a basic working knowledge of how data are stored in computer-readable form, and how data should be stored.
to acquire experience with the management, documentation and description of data in computer-readable form.
to acquire an appreciation of how data are used, to gain an appreciation of the technical and ethical problems surrounding the long-term storage and dissemination of data files

Content:

- major sources of data files
- 'dirty data'
- physical storage media
- access versus acquisition
- file structures and documentation
- elementary statistical analysis
- ethical considerations: citation, copyright, etc.
- institutional structures

Basic readings: students are expected to select their own reading material from an on-line SPIRES database

Method: in addition to lecture material, students will be assigned a data file which they will take through standard processing steps, generate appropriate documentation and cataloguing records, and finally do basic statistical analysis

Prerequisite: none

Evaluation: will be based on hands on work with an assigned data file and a take-home examination

Course credit: 1.5 units.

For further information, contact:
School of Library, Archival and Information Studies
University of British Columbia
2075 Wesbrook Mall
Vancouver, B.C. V6T 1W5
Phone: (604) 228-2404.

University of Alberta

Summer Institute on the
Quantitative Analysis of Social Data
June 20 – 30, 1988

Level I – Beginning Statistics

- basic descriptive statistics
- statistical inference
- regression analysis

Harvey Krahn
University of Alberta

Level II – Applied Regression Analysis

- build and test linear models
- interpret regression diagnostics
- logit analysis

Bob Arnold
Queen's University

Joint Sessions

Data Analysis Workshop

- discussion of data analysis problems
- stages of survey data analysis
- data interpretation

Bill Avison
University of Western Ontario

Computing & Data Analysis

- preparing data for statistical packages
- using packages to clean data
- labs using SPSSx and SAS

Chuck Humphrey & Wayne Watson
University of Alberta

Sponsors: Computing Services and the Population Research Laboratory

Cost: $650 per person (includes tuition, course materials and computing time)

Further information: write Ilze Hobin, Department of Sociology, University of Alberta, Edmonton, Alberta T6G 2H4 or phone (403) 432-4659.

IASSIST88

PRELIMINARY PROGRAM

IASSIST '88
Public Data: Use it or Lose it
Thursday, May 26, to Sunday, May 29, 1988
Dupont Plaza Hotel
Washington, DC
USA

WEDNESDAY
4:00 p.m. to 8:00 p.m.
REGISTRATION
DuPont Plaza Hotel, IASSIST Headquarters Suite
THURSDAY, MAY 26
8:00 a.m. to 5:00 p.m.
REGISTRATION
Lobby, DuPont Plaza Hotel

Morning sessions, 8:30 a.m. to 12:15 p.m.
Plenary session
Session 1: View from the National Statistical Agencies
Judith Rowe, Princeton University, Chair
William P. Butz, Associate Director, Demographic Fields Area, U.S. Bureau of the Census
Denis Desjardins, Director General, Marketing and Information Services, Statistics Canada
Edmund Rapaport, Assistant Director, Statistics Sweden
Concurrent Sessions
Session 2-A: The Preservation and Use of Empirical Economic Research Data
Dan Newlon, National Science Foundation
Carolyn Geda, ICPSR
Session 2-B: Educating the Data User: The Role of Bibliographic Instruction

Bliss B. Siman, Baruch College, City University of New York, Chair
Rona Ostrow, Baruch College, City University of New York
Kristin McDonough, Baruch College, City University of New York
Eleanor Langstaff, Baruch College, City University of New York
Bobbie Pollard, Baruch College, City University of New York
Ida Lowe, Baruch College, City University of New York

Session 2-C: Digital Cartographic Data
Randle W. Olsen, US Geological Survey
Robert LaMacchia, Bureau of the Census, "TIGER Files: Digital Cartography for the 1990 Census."

Afternoon Concurrent sessions, 1:30 p.m. to 5:00

Session 3-A: Data Analysis Using Microcomputers
Dianne Crispell, American Demographics Magazine, "The Use of Microcomputers for Demographic Analysis: An Overview of Options for the Novice User."
Ari Silva and Arthur Conning, UN Latin American Demographic Centre (Santiago, Chile), "Small-area Census Data Service by Microcomputer: Application of the REDATAM System in Latin America and the Caribbean."
Charles Post, Mathematica Policy Research, "The SIPP Electronic Codebook."
S. W. Weisman, National Technical Information Service, "Data Dissemination on Diskettes."

Session 3-B: Democratization of Data
John Gales, John Gales Associates
Betty Steiger, Information Resources Management Associates
Jack Taub, National Information Utilities

Session 3-C: Consumer Expenditure Survey
Eva Jacobs, Bureau of Labor Statistics
Stephanie Shipp, Bureau of Labor Statistics

Session 4-A: Storage Technology for Data Archives
Margaret O. Adams, National Archives and Records Administration, Chair
S. Nazim Ali, Bahrain University, "Use of CD-ROM in a University Library."
Laura Guy and Pat Wenzel, University of Wisconsin, "Data Storage on CD WORM."
S. Pal Asija, Attorney and Engineer, Shelton, Connecticut, "Handwritten Data Input, Storage Output, and Recognition System."

Session 4-B: Issues in the Integration of Machine-Readable Records into Traditional Library Service
Jaia Barrett, Association of Research Libraries, chair
Kathleen Heim, Louisiana State University
Karin Wittenborg, University of California at Los Angeles
Donald Treiman, University of California at Los Angeles
Raymond Carpenter, University of Nort Carolina at Chapel Hill

Session 4-C: Statistical Use of Administrative Records
Robert Dalrymple, U.S. Food and Nutrition Service

Carole Trippe, Mathematica Policy Research, Inc.
Michael Strudler and Peter Sailor, Internal Revenue Service

RECEPTION -- NATIONAL ARCHIVES, 5:30 p.m. - 7:30 p.m.

FRIDAY, MAY 27

8:00 a.m. to 5:00 p.m.
REGISTRATION
Lobby, DuPont Plaza Hotel

Morning sessions, 9:00 a.m. to 12:00 noon

Plenary session

Session 5: Professionalism and Professional Standards in the Information Age
Constance Citro, Committee on National Statistics, National Research Council, Chair and Moderator
Barbara Bailer, President, American Statistical Association
Frank B. Evans, President-elect, Society of American Archivists

Concurrent Sessions

Session 6-A: The Development and Management of Data Archives
Daniel C. Tsang, University of California -- Irvine, "Collection Development Policies for Machine-Readable Data Files."
Rudolph Bell, Rutgers University, "Medieval and Early Modern Data Bank."
Wong Poh Kam, Seres Sdn. Bhd. "Collection, Dissemination, and Preservation of New Science Data in Malaysia."
Martin Pawlocki and Elizabeth Stephenson, University of California at Los Angeles, "A Microcomputer Based Tape Information System."

Session 6-B: Data Management Technology
Derek Ballantyne, National Archives of Canada, "Using a Local Area Network as a Front End for Data Processing in an Archives."
Donald P. Trees, Viar & Company, "A Generalized Approach to Data Quality Assurance and Control of Large Complex Data Structures Using SAS."
David L. Sallach, University of Arkansas, "Knowledge Interchange Standards and th4e RM/T Model."

Session 6-C: Cross-National Social Surveys
Roger Jones, Australian National University, Moderator/Convenor
Speakers will discuss content and archival administration of the European Values Survey, the International Social Survey Project, the Class Structure and Class Consciousness survey, and other similar surveys.

ROUNDTABLE LUNCHES: 12:00 noon - 1:30 p.m.

Complex Data Sets — How to Make Them Friendly and usable — Convenor: Michal Peleg
Archival Appraisal of Machine-Readable Statistical Files — Convenor: Michael L. Miller, National Archives and Records Administration
Archival Tools — Convenors: Jacqueline McGee, The Rand Corporation
Complex Data Sets — Convenors: Pat Doyle, Mathematica Policy Research

and David McMillan, Bureau of the Census

Afternoon sessions 1:30 p.m. – 5:00 p.m.

Session 7–A, 8–A: Techniques and Tools of Microsimulation
William Bradley, Health and Welfare Canada, Chair
Steven Caldwell, Cornell University
Steve Gribble, Statistics Canada
Sharon Hirabayshi, Mathematica Policy Research, Inc.
Richard J. Morrison, Health and Welfare Canada
Randall Webb, Urban Institute
Bruce Winer, Carlton University
Michael Wolfson, Statistics Canada

Session 7–B: On-line Catalogues With Records of Computer Data Files
Lynn Marko, University of Michigan,
"Generalizing Access to Statistical Research Materials."
Ted Brandhorst, Director, ERIC Facility,
"Access to Description of Education-Related Machine-Readable Data Files Via the Bibliographic Data Base of the ERIC System."
Ed Hanis, Tycho Research, Associates,
"CULDAT: The Canadian Union List of Machine Readable Data Files"

Session 7–C: Editing Strategies
Kay Worrell, Conference Board, Chair
Speakers will discuss alternative strategies for editing survey and administrative data. Of interest is the effect of program or policy objectives on the choices for editing incomplete and inconsistent data.

Session 8–B: Issues in Data Dissemination
Katharine Sue Gavrel, National Archives of Canada, Chair
E. Walter Terry, Florida State University, "Adopting a Standard for Compression of Public Use Demographic Data."
Edie Hedlin, National Archives and Records Administration, "New Regulations on Access to Highly Personal Information for Statistical Research."

Session 8–C: Mapping as a Data Dissemination Technique
Michal Peleg, Social Science Data Archive, Hebrew University of Jerusalem,
"Presenting Spatial Data – The Statistical Map as a New Practice."
Alan Fox, Chadwyck-Healey, Inc.,
"Supermap: Revolutionizing the Way to Access and Manipulate Demographic Data."

SATURDAY, MAY 28

8:30 a.m. to 11:30 p.m.
REGISTRATION
Lobby, DuPont Plaza Hotel

Plenary session, 9:00 – 11:30

Session 9: Beyond the Social Science Data Archives
Hans Jorgen Marker, Danish Data Archives,
"History and the Data Archives."
Jake Th. V. Knoppers, Canadian Workplace Automation Research Center,
"The Archival Implications of Electronic Document/Data Interchange."
David Bearman, Archives and Museum Informatics,

"Collecting Software: A New Challenge for Archives and Museums."
Erwin Scheuch, University of Cologne Institute for Applied Social Research,
"Data as Part of a Social Science Infrastructure."
Felix Krayeski, Congressional Research Service,
"Storage and Dissemination Technology for Data Archives."

Lunch and Business Meeting, 11:30 a.m. – 1:30 p.m.

1:30 p.m. – 5:00 p.m. WORKSHOPS

(Requires separate registration)

Session 22: Interchange Standards (tentative)
Organizer: Ted Weir, National Archives and Records Administration

Session 23: Data Products From the Bureau of the Census:
1990 Decennial Data, Economic Census Data, Foreign Trade Data
Organizer: Barbara Aldrich, Bureau of the Census

Session 24: The Data Archive on Adolescent Pregnancy and
Pregnancy Prevention
Organizer: J. J. Card, Sociometrics Corporation

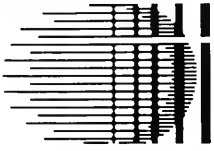
SUNDAY, MAY 29

9:30 a.m. – 12:30 p.m. WORKSHOPS

(Requires separate registration)

Session 25: Integrating Machine-Readable Records into
Traditional Library Services
Organizer: Elizabeth Stephenson, University of
California at Los Angeles

Session 26: Special Aspects of the National Logitudinal Surveys of Youth:
Focus on the NLSY Child-Mother Data
Organizer: Paula C. Baker, Center for Human Resource Research,
The Ohio State University



IASSIST

The International Association for Social Science Information Services and Technology (**IASSIST**) is an international association of individuals who are engaged in the acquistion, processing, maintenance, and distribution of machine readable text and /or numeric social science data. The membership includes information system specialists, data base librarians or administrators, archivists, researchers, programmers, and managers. Their range of interests encompases hard copy as well as machine readable data.

Paid-up members enjoy voting rights and receive the **IASSIST QUARTERLY**. They also benefit from reduced fees for attendance at regional and international conferences sponsored by **IASSIST.**

Membership fees are:
Regular Membership. $20.00 per calendar year.
Student Membership: $10.00 per calendar year.

Institutional subcriptions to the quarterly are available, but do not confer voting rights or other membership benefits.

Institutional Subcription: $35.00 per calendar year (includes one volume of the Quarterly)

Membership form

I would like to become a member of Iassist. Please see my choice below:

- $20 Regular Membership
- $10 Student Membership
- $35 Institutional Membership

My primary Interests are:

- Archive Services/ Administration
- Data Processing/Data Management
- Research Applications
- Other (specify) _____

Please make checks payable to IASSIST and Mail to :

Ms Jackie McGee
Treasurer, IASSIST
% Rand Corporation
1700 Main Street
Santa Monica
California USA 90406

Name/phone

Institutional Affiliation

Mailing Address

City

Country/zip/postal code

Editorial Information

The **IASSIST QUARTERLY** represents an international cooperative effort on the part of individuals managing, operating, or using machine-readable data archives, data libraries, and data services. The **QUARTERLY** reports on activities related to the production, acquisition, preservation, processing, distribution, and use of machine-readable data carried out by its members and others in the international social science community. Your contributions and suggestions for topics of interest are welcomed. The views set forth by authors of articles contained in this publication are not necessarily those of **IASSIST.**

Information for Authors

The **QUARTERLY** is published four times per year. Articles and other information should be typewritten and double-spaced. Each page of the manuscript should be numbered. The first page should contain the article title, author's name, affiliation, address to which correspondence may be sent, and telephone number. Footnotes and bibliographic citations should be consistent in style, preferably following a standard authority such as the University of Chicago press **Manual of Style** or Kate L. Turabian's **Manual for Writers**. Where appropriate, machine-readable data files should be cited with bibliographic citations consistent in style with Dodd, Sue A. "Bibliographic references for numeric social science data files: suggested guidelines". *Journal of the American Society for Information Science 30(2):77-82, March 1979.* If the contribution is an announcement of a conference, training session, or the like, the text should include a mailing address and a telephone number for the director of the event or for the organization sponsoring the event. Book notices and reviews should not exceed two double-spaced pages. Deadlines for submitting articles are six weeks before publication. Manuscripts should be sent in duplicate to the Editor:

Walter Piovesan, Research Data Library, W.A.C. Bennett Library, Simon Fraser University Burnaby, B.C., V5A 1S6 CANADA (01)604/291-4349 E-Mail: USERDLIB@SFU.BITNET

Book reviews should be submitted in duplicate to the Book Review Editor:

Daniel Tsang, Main Library, University of California P.O. Box 19557, Irvine, California 92713 USA. (01)714/856-4978 E-Mail: DTSANG@ORION.CF.UCI.EDU

Key Title: **Newsletter - International Association for Social Science Information Service and Technology**
ISSN - United States: 0739-1137

SANTA MONICA, California
U.S.A. 90406-2138

PRIORITY